

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS MATEMÁTICAS
Departamento de Estadística e Investigación Operativa



TESIS DOCTORAL

Contribuciones al análisis de los problemas de influencia y multicolinealidad en regresión lineal

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR

Santiago Velilla Cerdan

DIRECTOR:

Manuel del Rio Bueno

Madrid, 2015

TP
1988
123

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Ciencias Matemáticas

Departamento de Estadística e Investigación Operativa



x-53-162240-2

**CONTRIBUCIONES AL ANALISIS DE LOS
PROBLEMAS DE INFLUENCIA Y
MULTICOLINEALIDAD EN REGRESION LINEAL**

Santiago Velilla Cerdán

Madrid, 1988

Colección Tesis Doctorales. N.º 123/88

© **Santiago Velilla Cerdán**

Edita e imprime la Editorial de la Universidad
Complutense de Madrid. Servicio de Reprografía
Noviciado, 3 - 28015 Madrid
Madrid, 1988
Ricoh 3700
Depósito Legal: M-4567-1988

CONTRIBUCIONES AL ANALISIS
DE LOS PROBLEMAS
DE INFLUENCIA Y MULTICOLINEALIDAD
EN REGRESION LINEAL

Memoria presentada para optar al
grado de Doctor en Ciencias Mate-
máticas por Santiago Velilla Cer-
dán y dirigida por el Dr.D.Manual
del Río Bueno.

a mis padres

<u>INDICE</u>	Pág.
PROLOGO	2
CAPITULO I: <u>Introducción</u>	4
1. Sumario	4
2. Descripción de los elementos básicos	4
3. Influencia de los casos	7
3.1. Planteamiento del problema	7
3.2. Medidas de influencia	9
3.2.1. Distancia de Cook	10
3.2.2. Estadístico de Andrews-Pregibon	11
3.2.3. Medidas de Belsley, Kuh y Welsch	12
3.2.4. Cambio de volumen de un elipsoide de confianza	13
3.3. Ejemplo	14
4. Multicolinealidad en regresión lineal	16
4.1. Definición	16
4.2. Causas y efectos	19
4.3. Diagnósticos de multicolinealidad	23
4.3.1. Coeficientes de correlación	23
4.3.2. Factores de inflación de la varianza.	24
4.3.3. Módulo de Gunst	24
4.3.4. Autosistema de la matriz de corre- lación	26
4.3.5. Diagnósticos de BKW	26
4.4. Ejemplo	27
5. La matriz ampliada X^*	29
CAPITULO II: <u>LOS DIAGNOSTICOS DE MULTICOLINEALIDAD EN LA MATRIZ AMPLIADA</u>	
1. Sumario	31
2. Fórmulas básicas. Introducción	31
3. Análisis de los VIF_j^* ($j=1,2,\dots,m$)	34
3.1. Detección de multicolinealidad en Z con los VIF_j^* ($j=1,2,\dots,p$)	35
3.1.1. Método directo	35
3.1.2. Distribución de los VIF_j^* ($j=1,\dots,p$).	36

4. Autosistema de la matriz de productos cruzados	
Z^*Z^*	37
4.1. Autovalores	37
4.1.1. Teorema de separación. Consecuencias .	37
4.1.2. Distribución de los autovalores	40
4.2. Autovectores	42
4.3. Desigualdades de Berk	43
5. Módulo de Gunst	44
5.1. Relación entre los módulos de Z y Z^*	44
5.2. Comentarios	46
5.3. Ejemplo	48
6. Ejemplo	50

CAPITULO III: UN ANALISIS CONJUNTO DE MULTICOLINEALIDAD E INFLUENCIA

1. Sumario	55
2. Motivación	55
3. Análisis de multicolinealidad en Z_1	57
3.1. Factores de inflación de la varianza	57
3.2. Autovalores e índices de condición	58
3.3. Módulo de Gunst	61
3.4. Desigualdades de Berk	64
3.5. Aplicaciones: detección simultánea de colinealidad y potencial	65
3.6. Ejemplo	65
3.7. Desarrollos sin tipificar los datos	69
4. La matriz Z_1^*	71
4.1. Inversa de $Z_1^*Z_1^*$	71
4.2. Diagnósticos de condición en Z_1^*	74
4.2.1. Factores de inflación de la varianza .	74
4.2.2. Otros resultados	75
4.3. Detección de multicolinealidad y casos influyentes	76
4.4. Ejemplo	75
4.5. Datos sin tipificar	83

CAPITULO IV: INFLUENCIA DE LOS CASOS EN LOS DIAGNOSTICOS
DE MULTICOLINEALIDAD

1. Sumario	05
2. Motivación	06
3. Diagnósticos como funcionales	07
3.1. Diferenciabilidad de $D^* = I [F_n^*]$	08
3.2. Fórmula para $D^*-D_{(i)}^*$	09
3.3. Calidad y mejora de la aproximación	90
3.4. Ejemplo	91
4. Influencia en el autosistema de la matriz de productos cruzados	92
4.1. Introducción	92
4.2. Influencia en los autovalores	93
4.2.1. Autovalores como funcionales	93
4.2.1.a) Datos sin tipificar	93
4.2.1.b) Datos tipificados	94
4.2.2. Aplicaciones	95
4.2.2.a) Interpretación de resultados	95
4.2.2.b) Comportamiento de los gráficos de componentes principales	101
4.2.2.c) Influencia en componentes prin- cipales	107
4.2.3. Otras técnicas	107
4.2.3.a) Datos sin tipificar	107
4.2.3.b) Datos tipificados	108
4.3. Influencia en los autovectores	111
4.3.1. Autovectores como funcionales	111
4.3.2. Caso particular (Criterio de Wilks) ..	113
5. Influencia en los factores de inflación de la varianza	114
5.1. Método directo	114
5.1.1. Datos sin tipificar	114
5.1.2. Datos tipificados	117
5.2. Los VIF_j^* como funcionales	117
6. Otros diagnósticos	120
6.1. Coeficientes de correlación	120
6.2. Determinante de $Z'Z$	122

APENDICE

A.1. Fórmulas matriciales	125
A.1.1. Fórmulas de actualización	125
A.1.2. Inversa de una matriz en cajas	125
A.2. Diferenciación de funcionales en espacios lineales	125
A.2.1. Definiciones y consecuencias	125
A.2.2. Derivadas de orden superior	127
A.2.3. Algunos resultados útiles	127
A.3. Teoría de perturbaciones para matrices simétricas	130
A.3.1. $C[\alpha] = A + \alpha B$	130
A.3.2. $C = A + B$	132
BIBLIOGRAFIA	133

CONTRIBUCIONES AL ANALISIS
DE LOS PROBLEMAS
DE INFLUENCIA Y MULTICOLINEALIDAD
EN REGRESION LINEAL

PROLOGO

El ajuste de un modelo lineal de regresión por mínimos cuadrados es una herramienta estadística de aplicación universal. La relativa sencillez rutinaria del método está lastrada por gran variedad de inconvenientes. Dos de los mismos destacan con luz propia: el problema de las observaciones o casos influyentes y el problema de la multicolinealidad o de dependencias lineales aproximadas entre las variables regresoras del modelo.

De modo general, el objetivo de este trabajo es aportar varios nuevos métodos de enfoque y tratamiento de estas dos últimas cuestiones. Los resultados obtenidos están expuestos en los capítulos II, III y IV, mientras que el Capítulo I es una introducción, destinada a enmarcar la investigación y fijar las notaciones y terminología. Hay que observar que los capítulos II al IV están, en cierto sentido relacionados entre sí. El contexto de los desarrollos es, esencialmente, aplicado y las nuevas ideas presentadas se complementan con análisis de ciertos conjuntos de datos. Determinados resultados teóricos, necesarios en la exposición, están recogidos en un apéndice adicional.

Como queda dicho, el Capítulo I es una introducción que contiene la presentación de los problemas de influencia y multicolinealidad y la fijación del lenguaje empleado. La bibliografía disponible es muy numerosa y puede encontrarse brevemente reseñada en este capítulo. Puesto que uno de los aspectos originales de este trabajo se refiere a cierta utilización de la matriz de diseño aumentada, este capítulo incluye un apartado dedicado a la misma.

El capítulo II es un estudio del problema de multicolinealidad en la matriz ampliada. En concreto se obtienen varios resultados y su relación con la detección de multicolinealidad en la matriz de diseño original.

En el Capítulo III, se expone un método nuevo de análisis combinado de los problemas de influencia y multicolinealidad. La idea básica es introducir una ligera perturbación en la matriz de

diseño, aumentada o no, que tratada con elementos propuestos originalmente como técnicas de análisis de condición proporciona información simultánea sobre ambos problemas.

Finalmente, el Capítulo IV contiene un estudio de la influencia de casos individuales en los diagnósticos de multicolinealidad más usuales. Destaca el desarrollo de una técnica basada en propiedades de diferenciabilidad de funcionales como herramienta necesaria en los desarrollos de este capítulo.

Después del apéndice, el trabajo termina con un apartado dedicado a recoger las referencias bibliográficas utilizadas.

Esta visión general del trabajo, puede completarse con los sumarios incluidos al principio de los capítulos que contienen resúmenes detallados de la materia tratada en los mismos.

En el turno de agradecimientos, quisiera citar en primer lugar al profesor Manuel del Río por su dirección y ayuda constante. Los profesores Antonio Cuevas y Juan Romo han colaborado en la técnica relativa a la diferenciación de ~~funcionales~~ del capítulo IV. Varios profesores del Departamento, han intervenido en el tratamiento informático de los ejemplos. Además de estos nombres, hay otras personas que se han interesado de manera constante por la evolución de este trabajo. A todos ellos, mi más sincera gratitud.

Madrid, Noviembre de 1986

CAPITULO I

1. SUMARIO

Este capítulo está dedicado a la descripción de los elementos básicos sobre los que descansa la investigación desarrollada en los Capítulos II, III y IV.

Así, en la sección 2 se encuentra, por un lado, un esquema de los resultados más importantes del análisis mínimo-cuadrático de un modelo lineal de regresión y, por otro, el establecimiento de la notación empleada en varias magnitudes y estadísticos de frecuente referencia en este trabajo.

La sección 3 es un amplio resumen del problema de influencia en modelos lineales. Tras una breve presentación del mismo, se describen algunas de las medidas de influencia más aceptadas. Un ejemplo muy conocido es el complemento final de esta sección.

La sección 4 contiene aspectos esenciales del problema de multicolinealidad en modelos lineales. Tras la definición, se reseñan las causas y efectos de una dependencia lineal aproximada entre las variables del modelo. Se hace especial hincapié en los diagnósticos de multicolinealidad más utilizados en la práctica. Esta sección termina asimismo con un sencillo ejemplo.

Finalmente, la sección 5 es la presentación de la matriz ampliada asociada a un modelo lineal que juega un papel central en este trabajo. Por razones de completitud,

se incluye la demostración de un resultado relativo a la caracterización de los elementos diagonales de la matriz de proyección ampliada como medidas de la influencia de un caso.

2. DESCRIPCION DE LOS ELEMENTOS BASICOS

El marco de referencia fundamental es el modelo lineal de regresión múltiple

$$Y = X\beta + \varepsilon \quad (1.1)$$

en el que $Y = (Y_1, \dots, Y_n)'$ es un vector $n \times 1$ de observaciones y X es una matriz $n \times m$ de rango completo que se supondrá siempre con término independiente, es decir $X = [1_n | X_1]$, donde X_1 es $n \times p$ y, por tanto, $m = p + 1$. β es un vector $m \times 1$ de parámetros desconocidos y σ^2 un factor de escala positivo. De acuerdo con la partición de X puede considerarse que $\beta = (\beta_0 | \beta_1)'$, (β_1 es de $p \times 1$). El término de error ϵ es un vector aleatorio n -dimensional con momentos $E[\epsilon] = 0$ y $V[\epsilon] = \sigma^2 I_n$. Aunque no es imprescindible, siempre que sea necesario puede suponerse la hipótesis usual de normalidad $\epsilon \sim N_n(0; \sigma^2 I_n)$.

Los resultados del análisis del modelo (1.1) por mínimos cuadrados (m.c.) son suficientemente conocidos y sólo será necesario, para una utilización posterior, fijar la notación empleada. Así, se escribirá

$$\hat{\beta} = (X'X)^{-1} X'Y \quad = \text{estimador de m.c. de } \beta$$

$$\hat{Y} = X\hat{\beta} \quad = \text{vector de valores ajustados.}$$

$$C[X] = \text{variedad lineal engendrada por las columnas de la matriz } X.$$

$$V = (\sigma_{ij}) = X(X'X)^{-1}X' \quad = \text{matriz de proyección sobre } C[X]$$

$$e = Y - X\hat{\beta} = (I - V)Y \quad = \text{vector de residuales m.c.}$$

$$SCE = e'e = \sum_{i=1}^n e_i^2 \quad = \text{suma de cuadrados de los errores.}$$

$$\hat{\sigma}^2 = \frac{SCE}{n-m} \quad = \text{estimador centrado de } \sigma^2$$

También será necesaria la descomposición de análisis de la varianza en sus dos versiones

$$SCT = SCR + SCE \quad (1.2)$$

y

$$SCT_m = S^2_y = SCR_m + SCE \quad (1.3)$$

siendo

$$SCT = \sum_{i=1}^n Y_i^2 \quad = \text{Suma de cuadrados total}$$

$$\begin{aligned} SCR &= \hat{\beta}' X' Y = Y' V Y = \text{Suma de cuadrados debida a la regresión} \\ SCT_m &= S_y^2 = \sum_{i=1}^n (y_i - \bar{y})^2, \left(\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \right) = \text{Suma de cuadrados corregida por la media} \end{aligned}$$

$$SCR_m = SCT_m - SCE = \text{Suma de cuadrados debida a la regresión corregida por la media.}$$

Es sabido que la inclusión del término independiente en la matrix X permite reescribir el modelo (1.1) en la forma

$$Y = 1_n \beta_0 + X_1 \beta_1 + \varepsilon = 1_n \delta_0 + \mathbb{X} \beta_1 + \varepsilon = 1_n \alpha_0 + Z \alpha + \varepsilon$$

donde \mathbb{X} es la matrix $n \times p$ de datos centrados y Z es la matrix $n \times p$ de datos tipificados. La matrix $Z'Z$ es, pues, la matrix de correlaciones asociada a (1.1). También se tiene, en virtud del proceso de ortogonalización

$$\begin{aligned} V &= X (X'X)^{-1} X' = \frac{1_n 1_n'}{n} + \mathbb{X} (\mathbb{X}'\mathbb{X})^{-1} \mathbb{X}' = \\ &= \frac{1_n 1_n'}{n} + Z (Z'Z)^{-1} Z' \end{aligned} \quad (1.4)$$

Estas dos transformaciones, la centralización y la tipificación, son fundamentales en el tratamiento de modelos lineales.

Una herramienta de uso constante en este trabajo es la descomposición espectral de las matrices productos cruzados $X'X$ y $Z'Z$. Con el fin de simplificar la notación, ésta descomposición se escribirá siempre en la forma CDC' donde C es una matrix ortogonal cuyas columnas son autovectores de la matrix correspondiente y $D = \text{diag}(\lambda_i)$ es una matrix diagonal de elementos diagonales los autovalores asociados. Dependiendo del contexto, C será de orden $m \times m$ o $p \times p$, ..., etc. Se adopta como criterio de ordenación de los autovalores el sentido creciente $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m$ (o $\lambda_1 \leq \dots \leq \lambda_p$, en el caso de la matrix $Z'Z$).

Se denotará por x_i' una fila genérica de la matrix X y por caso se entenderá la estructura conjunta (x_i', y_i) . De acuerdo con lo exigido a la matrix X es $x_i' = (1, x_{i1}')$ donde x_{i1} es $p \times 1$. En relación con un caso, es importante la magnitud σ_{ii} , i -ésimo elemento diagonal de la matrix V , que puede escribirse

$$\begin{aligned} \sigma_{ii} &= x_i' (X'X)^{-1} x_i = \frac{1}{n} + (x_{i1} - \bar{x})' (\mathbb{X}'\mathbb{X})^{-1} (x_{i1} - \bar{x}) = \\ &= \frac{1}{n} + z_i' (Z'Z)^{-1} z_i = \frac{1}{n} + \sigma_{ii} [Z] \end{aligned} \quad (1.5)$$

en virtud de (1.4), \bar{x} es el centroide de las filas de X_1 y \bar{z}_i es la i -ésima fila de la matriz Z . Asociados a un caso están también e_i = i -ésimo residual m.c. y, dado que, $\text{var}[e_i] = \sigma^2(1 - \sigma_{ii})$, los residuales estudentizados

$$r_i = \frac{e_i}{\hat{\sigma} \sqrt{1 - \sigma_{ii}}} = \text{i-ésimo residual estudentizado interno}$$

y

$$t_i = \frac{e_i}{\hat{\sigma}_{(i)} \sqrt{1 - \sigma_{ii}}} = \text{i-ésimo residual estudentizado externo} \quad (1.6)$$

A lo largo del trabajo, la notación (i) indica que la magnitud correspondiente está calculada sin considerar el caso i -ésimo. En virtud de la fórmula

$$SCE = SCE_{(i)} + \frac{e_i^2}{1 - \sigma_{ii}}$$

los residuales r_i y t_i de (1.6) están relacionados de forma monótona (Cook y Weisberg, (1982) pág. 19).

Finalmente, a lo largo de la exposición será necesario utilizar de manera repetida ciertos resultados de inversión de matrices y cálculo de determinantes. Todos ellos se encuentran en el apéndice. En este último se incluye también un breve resumen de la teoría de perturbaciones para matrices simétricas, de utilidad en el Capítulo IV.

3. INFLUENCIA DE LOS CASOS

3.1. Planteamiento del problema

Es conocido que el estimador de mínimos cuadrados $\hat{\beta}$ más puede reflejar características peculiares de un caso o reducido grupo de casos que describir con precisión la relación lineal entre la respuesta Y y las variables independientes X . Por ejemplo, en regresión polinómica un determinado coeficiente $\hat{\beta}_j$ puede ser significativo debido únicamente a la presencia de un cierto caso en el conjunto de datos (Cook-Weisberg, 1982, pág. 111).

En un análisis por mínimos cuadrados, un caso se dice influyen

te cuando su eliminación del conjunto de datos altera sustancialmente los resultados (estimador $\hat{\beta}$, valores ajustados \hat{Y} , ... etc). Desde un punto de vista general, Weisberg (1983) señala que la idea general de un análisis de influencia es estudiar los cambios provocados en ciertos aspectos del análisis de un modelo cuando se aplica un determinado esquema de perturbación a los datos. Según esta idea, la influencia de un caso en el modelo $Y = X\beta + \varepsilon$ se incluye dentro del problema general de influencia considerando el esquema de perturbación de eliminación de casos.

La bibliografía del problema de influencia en regresión es muy amplia. Los artículos de Cook (1977, 1979) son puntos fundamentales de la teoría, continuados después por Cook y Weisberg (1980) y el libro básico de referencia de Cook y Weisberg (1982). (CW). Es interesante también el libro de Belsley, Kuh y Welsch (1980) (BKW). Uno de los trabajos más citados es debido a Andrews y Pregibon (1978). Draper y John (1981) estudian y comparan los resultados de Cook y Andrews y Pregibon. Atkinson (1982) sugiere varias ideas nuevas, en especial con el problema de influencia no en el estimador sino en relación con el parámetro de Box-Cox. Weisberg (1983) propone varios principios básicos para efectuar un análisis de influencias en modelos lineales. Problemas relacionados con el parámetro de Box-Cox reciben tratamiento detallado en un artículo de Cook y Wang (1983). Otros trabajos relacionados se deben a Johnson y Geisser (1982, 1983) quienes, desde un punto de vista bayesiano, estudian el problema de la influencia en relación con la predicción. Gray y Ling (1985) desarrollan un método de detección de casos influyentes basado en la técnica de "clustering". Recientemente, Cook (1986) ha iniciado los primeros pasos en el concepto de la llamada influencia local que es, en pocas palabras, el análisis, mediante la distancia de verosimilitud, del efecto que producen pequeñas desviaciones en las especificaciones del modelo (1.1). Atkinson (1986) vuelve de nuevo, a problemas relacionados con las transformaciones.

Por lo que se refiere al presente trabajo, siempre que aparezca la palabra influencia, ésta se referirá a la influencia de un caso o grupo de casos relativa al esquema de perturbación de eliminación de casos.

3.2. Medidas de influencia

La investigación del problema se ha centrado en el desarrollo de medidas de la influencia de un caso (x_i', Y_i) . Básicamente, una medida es una función de los datos que permite calibrar o evaluar la influencia de un caso en un determinado aspecto del análisis. Una vez que una medida detecta un caso influyente, éste debe examinarse detalladamente para averiguar, en lo posible, las razones de tal peculiaridad. Si esta última puede atribuirse a errores de medida o registro (p.ej. un porcentaje mayor que 100%) el caso debe eliminarse del análisis, o si ello fuera posible, modificarse con sus valores correctos. La eliminación del caso es, en general, bastante discutible. Una recomendación de carácter general es no tomar ninguna decisión a este respecto de manera mecánica o rutinaria.

Cada caso posee dos componentes que pueden convertirlo en influyente: la parte aleatoria Y_i y la parte no estocástica x_i' . La respuesta Y_i puede tener un valor anormalmente grande o pequeño aún cuando corresponda a unos valores en $x_i' = (1, x_{i1}')$ de carácter razonable. Tales casos son los llamados "outliers". Un "outlier" puede detectarse con un valor de e_i extremo o, para evitar problemas de varianza no constante, con un valor aberrante de r_i o t_i . De hecho, t_i^1 es el estadístico del test de razón de verosimilitudes (o test-F) para la hipótesis $H_0: \phi = 0$ frente a $H_1: \phi \neq 0$ en el modelo de media trasladada

$$Y = X\beta + d_i\phi + \varepsilon, \quad \varepsilon \sim N_n(0; \sigma^2 I_n)$$

, $d_i = (0, \dots, 1^{(i)}, \dots, 0)'$, herramienta usual en el tratamiento de "outliers". Por otro lado, la fila $x_i' = (1, x_{i1}')$ de la matriz X puede ser extrema en el sentido de que x_{i1}' , como punto de \mathbb{R}^p , está alejado del resto de los datos x_{j1}' ($j \neq i$). Una medida de la lejanía de x_i' al centro de los datos es el valor de v_{ii} que, según la representación de (1.5),

$$v_{ii} = \frac{1}{n} + (x_{i1} - \bar{x})' (X'X)^{-1} (x_{i1} - \bar{x})$$

es una distancia de x_{i1} al centroide \bar{x} dada por la matriz de producto escalar $(X'X)^{-1}$. Se sabe que, en presencia de término independiente, es $\frac{1}{n} \leq v_{ii} \leq 1$. Valores cercanos de v_{ii} a 1 son

indicativos de puntos lejanos en el espacio de variables regresadas. En ocasiones, la terminología se confunde y se habla indistintamente de casos, observaciones y filas "outliers". En ocasiones, se hablará de v_{ii} como el potencial del caso i -ésimo. El primer trabajo relacionado con los v_{ii} se debe a Hoaglin y Welsch (1978).

Como se ilustra a continuación, las medidas de influencia existentes combinan información del bloque básico (v_{ii} , v_{ii}) o, equivalentemente, (t_i, v_{ii}) .

3.2.1. Distancia de Cook

Cook (1977) propone el estadístico más utilizado para evaluar la contribución o influencia de un caso en la determinación de $\hat{\beta}$, el estimador m.c. de β . La idea básica es construir una medida que refleje la distancia de $\hat{\beta}$ a $\hat{\beta}_{(i)}$, el estimador calculado eliminando el caso i -ésimo. Así, la distancia de Cook es el estadístico

$$D_i = \frac{[\hat{\beta} - \hat{\beta}_{(i)}]' X'X [\hat{\beta} - \hat{\beta}_{(i)}]}{m \hat{\sigma}^2} \quad (1.7)$$

Esta medida es la distancia al cuadrado de $\hat{\beta}$ a $\hat{\beta}_{(i)}$ dada por la matriz de producto escalar $X'X$ y tipificada por el factor de escala $m \hat{\sigma}^2$.

Un caso se juzga influyente si produce un valor grande de la distancia D_i . La magnitud de (1.7) se puede evaluar comparando con una tabla de valores de la distribución $F_{m, n-m}$ usando la semejanza entre D_i y el elipsoide de confianza conjunto de coeficiente $1-\alpha$, formado por aquellos puntos β^* tales que

$$\frac{(\beta^* - \beta)' X'X (\beta^* - \beta)}{m \hat{\sigma}^2} \leq F_{m, n-m, \alpha}$$

Si, por ejemplo, $D_i \approx F_{m, n-m, \alpha}$, la supresión del caso i -ésimo traslada el estimador hasta la frontera de un elipsoide de confianza $1-\alpha$ para β . Sin embargo, D_i no se distribuye como una F .

La distancia (1.7) se puede calcular con la información disponible al efectuar un análisis con todos los datos porque se tiene

$$D_i = \frac{1}{m} r_i^2 \frac{\sigma_{ii}}{1 - \sigma_{ii}} \quad (1.8)$$

La expresión (1.8) factoriza D_i , aparte del término constante m^{-1} , en el producto de r_i^2 y $\frac{\sigma_{ii}}{1 - \sigma_{ii}}$, función monótona de la distancia de x_{ii} al centro de los datos. Por consiguiente, según el estadístico de Cook, un caso puede ser influyente si la observación Y_i es "outlier" o si la fila x_i' es extrema.

La medida (1.7) es un caso particular de una clase de medidas dada por

$$D_i[M, c] = \frac{[\hat{\beta} - \hat{\beta}_{(i)}]' M [\hat{\beta} - \hat{\beta}_{(i)}]}{c} \quad (1.9)$$

donde c es un factor de escala positivo y M una matriz definida positiva. Puede demostrarse que la CIM (Curva de influencia muestral) del estimador $\hat{\beta}$ es $CIM = (n-1)[\hat{\beta} - \hat{\beta}_{(i)}]$ de modo que (1.9) es una norma o aplicación de \mathbb{R}^m en \mathbb{R}^+ de la CIM para ordenar los casos según su influencia. La elección de M y c y detalles elaborados de la teoría están muy bien explicados en el libro de Cook y Weisberg (1982).

3.2.2. Estadístico de Andrews-Pregibon

Otro método de detección de casos influyentes está dado por Andrews y Pregibon (1978). Sea $X^* = [X | Y]$ la matriz de constantes del modelo aumentada con la columna de observaciones Y . El estadístico de Andrews y Pregibon para evaluar la influencia del caso i -ésimo es

$$R_i = \frac{|X_{(i)}^{*'} X_{(i)}^*|}{|X^{*'} X^*|} \quad (1.10)$$

Geométricamente, $1 - R_i^{1/2}$ es la proporción de volumen de un elipsoide engendrado por $X^{*'} X^*$ atribuible al caso i -ésimo. La expresión (1.10) se puede factorizar en la forma

$$R_i = \left[1 - \frac{r_i^2}{n-m} \right] [1 - \sigma_{ii}]$$

que, de nuevo, combina información basada en el bloque (n, σ_u) . Un caso se considera influyente cuando R_i es pequeño. Como método para calibrar la magnitud del $\min R_i$, Andrews y Pregibon sugieren un método basado en el p-valor

$$p = P_r \{ R. \leq R.^{\circ} \}$$

donde $R_i = \min R_i$ y $R.^{\circ}$ es el valor observado de $R.$.

La idea inicial de Andrews y Pregibon se refiere a grupos de casos $I = [i_1, \dots, i_k]$ en lugar de a casos individuales. El análogo R_I de (1.10) factoriza en la forma

$$R_I = \left[1 - \frac{r_I^2}{n-m} \right] |I_k - V_I| \quad (1.11)$$

donde r_I^2 es un análogo para casos múltiples del residual estudentizado r_i^2 y el determinante $|I_k - V_I|$ depende de la caja $k \times k$ de V formada por las filas y columnas contenidas en I . La determinación de R_I para todos los subconjuntos de k casos puede ser complicada lo que constituye un inconveniente potencial de la medida (1.11). Recientemente, Little (1985) ha estudiado con mayor detalle el término $\frac{r_I^2}{n-m}$ de (1.11).

3.2.3. Medidas de Belsley, Kuh y Welsch

Belsley, Kuh y Welsch (1980) introducen los estadísticos DF y la técnica COVRATIO.

Los estadísticos DF se llaman así porque dependen de diferencias de resultados de un análisis calculados con todos los datos y con el i -ésimo caso suprimido. Cuando el interés principal radica en el estimador, la cantidad pertinente es

$$DFBETA_i = \hat{\beta} - \hat{\beta}_{(i)} \quad (1.12)$$

Si $C = (X'X)^{-1}X'$, se tiene $\text{var}[\hat{\beta}_j] = \sigma^2 \sum_{k=1}^n c_{jk}^2$, ($j = 1, \dots, m$) y, por consiguiente, una posible tipificación de (1.12), componente a componente, es

$$DFBETAS_{ij} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sigma_{(i)}^2 \left[\sum_{k=1}^n c_{jk}^2 \right]^{1/2}} = \frac{c_{ij}}{\left[\sum_{k=1}^n c_{jk}^2 \right]^{1/2}} t_i \quad (1.13)$$

$$i = 1, \dots, n, \quad j = 1, \dots, m$$

Cuando el objetivo es buscar casos influyentes en las predicciones, los estadísticos DF apropiados son

$$\text{DIFFIT}_i = \hat{Y}_i - \hat{Y}_{i(i)}$$

donde $\hat{Y}_i = x_i' \hat{\beta}$ e $\hat{Y}_{i(i)} = x_{i(i)}' \hat{\beta}_{(i)}$. Puesto que $\text{var}[\hat{Y}_i] = \sigma^2 \sigma_{ii}$ es natural tomar $\hat{\sigma}_{(i)}(\sigma_{ii})^{1/2}$ como factor de escala y considerar

$$\text{DIFFITS}_i = \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{\hat{\sigma}_{(i)}(\sigma_{ii})^{1/2}} = \left(\frac{\sigma_{ii}}{1 - \sigma_{ii}} \right)^{1/2} t_i \quad (1.14)$$

La introducción de $\hat{\sigma}_{(i)}$ en el denominador de (1.13) y (1.14) implica la independencia de éste con el numerador correspondiente.

La técnica COVRATIO está diseñada para comparar las matrices de covarianzas de $\hat{\beta}$ y $\hat{\beta}_{(i)}$. Teniendo en cuenta que $\text{var}[\hat{\beta}] = \sigma^2 (X'X)^{-1}$ y $\text{var}[\hat{\beta}_{(i)}] = \sigma^2 (X'_{(i)} X_{(i)})^{-1}$, el estadístico es

$$\begin{aligned} \text{COVRATIO} &= \frac{|\hat{\sigma}_{(i)}^2 (X'_{(i)} X_{(i)})^{-1}|}{|\hat{\sigma}^2 (X'X)^{-1}|} = \\ &= \left[\frac{n-(m+1)}{n-m} + \frac{t_i^2}{n-m} \right]^{-m} (1 - \sigma_{ii})^{-1} \end{aligned} \quad (1.15)$$

Los criterios propuestos para decidir cuándo un caso es influyente son $|\text{DFBETAS}_{i,j}| > 2/\sqrt{n}$, $|\text{DIFFITS}_i| > 2\sqrt{p/n}$ y $|\text{COVRATIO} - 1| > 3p/n$

3.2.4. Cambio de volumen de un elipsoide de confianza

Si A es una matriz $k \times k$ definida positiva y E es el elipsoide definido por $E = \{z \in \mathbb{R}^k : z'Az \leq a\}$, su volumen es $\text{vol}[E] = \frac{(a\pi)^{k/2}}{\Gamma(k/2+1)} |A|^{-1/2}$. Por consiguiente, si E_0 es el elipsoide de confianza conjunto para β de coeficiente $1-\alpha$ basado en el conjunto total de datos y $E_{(i)}$ es este mismo elipsoide, pero con el caso i -ésimo eliminado, el logaritmo del cociente o razón de volúmenes es

$$\begin{aligned} \log \frac{\text{vol}[E_0]}{\text{vol}[E_{(i)}]} &= \frac{1}{2} \log \left[\left(\frac{\hat{\sigma}^2}{\hat{\sigma}_{(i)}^2} \cdot \frac{F_\alpha}{F_{(i), \alpha}} \right)^{-m} \frac{|X'X|^{-1}}{|X'_{(i)} X_{(i)}|^{-1}} \right] \\ &= \frac{m}{2} \log \frac{F_\alpha}{F_{(i), \alpha}} - \frac{1}{2} \log \text{COVRATIO} \end{aligned} \quad (1.16)$$

de donde $\text{vol}[E_0] / \text{vol}[E_{(i)}] = \text{const} (\text{COVRATIO})^{-1/2}$. Por consiguiente, el cambio en el volumen es equivalente al estadístico COVRATIO. En (1.16), $F_{\alpha} = F_{m, n-m, \alpha}$ y $F_{(i), \alpha} = F_{m, n-(m+1), \alpha}$.

Si la cantidad (1.16) es grande y positiva, la eliminación del caso i -ésimo producirá una disminución sustancial del volumen. Teniendo en cuenta la expresión (1.15) para COVRATIO, un "outlier" (t_i^2 grande) puede producir este resultado. Si $\sigma_i^2 \approx 1$, la eliminación del caso puede causar un aumento sustancial de volumen. La mayor parte de los casos debe tener un cociente (1.16) próximo a cero.

3.3. Ejemplo

Un ejemplo puede servir de ayuda para fijar las ideas de este apartado. Considérese el modelo lineal de regresión simple $Y = \beta_0 + \beta_1 x + \epsilon$ para el conjunto de datos de la tabla 1.1. Estos datos proceden originalmente de Mickey, Dunn y Clark (1967) y han sido analizados en Andrews-Pregibon (1978), Draper y John (1981) y Cook-Weisberg (1982), entre otros. Las variables son x = edad de un niño al hablar (en meses) e y = puntuación en un determinado examen.

TABLA 1.1

Caso	x	y	Caso	x	y
1	15	95	11	7	113
2	26	71	12	9	96
3	10	83	13	10	83
4	9	91	14	11	84
5	15	102	15	11	102
6	20	87	16	10	100
7	18	93	17	12	105
8	11	100	18	42	57
9	8	104	19	17	121
10	20	94	20	11	86
			21	10	100

La nube de puntos asociada a estos datos se recoge en la Fig. 1.1.

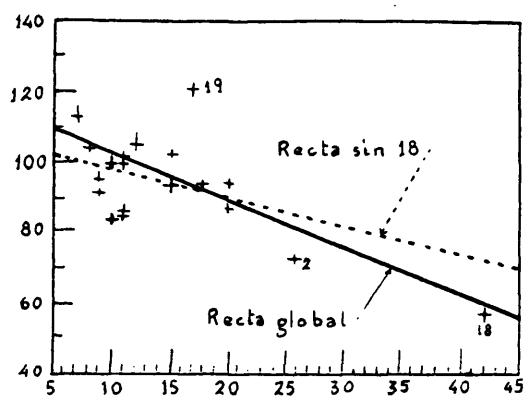


Fig. 1.1

Los puntos 2, 18 y 19 parecen dominar la tendencia lineal de los datos. Si estos tres puntos se eliminaran la sensación de linealidad sería más débil.

Considérense los puntos 18 y 19 con más detalle. Las rectas ajustadas con todos los datos, eliminando el caso 18 y eliminando el caso 19 son, respectivamente,

Conjunto total	$\hat{Y} = 109.87 - 1.13 x$
Caso 18 eliminado	$\hat{Y} = 105.63 - 0.78 x$
Caso 19 eliminado	$\hat{Y} = 109.30 - 1.19 x$

Como ya sugería un simple razonamiento sobre el dibujo, el caso 19 apenas influye en el ajuste, mientras que la eliminación del caso 18, aunque no altera esencialmente el término independiente, incide críticamente en la pendiente de la recta reduciéndola, en módulo, en un 30%.

El párrafo anterior queda refrendado por los valores de las medidas de influencia más usuales. La tabla 1.2 muestra los valores de ϕ_{ii} , r_i^2 , D_i y R_i para los casos 2, 18 y 19.

TABLA 1.2.

Caso	r_i^2	σ_{ii}	D_i	R_i
2	0.89	0.155	0.08	0.81
18	0.73	0.652	0.68 (max D_i)	0.33 (min R_i)
19	7.97	0.053	0.22	0.55

Con esta información se deduce que el caso 19 es un "outlier" como muestra el alto valor de r_i^2 . El p-valor para el contraste de la hipótesis $H_0: \phi = 0$ en el modelo de media trasladada para este caso es menor o igual que 0.0425. La anomalía del caso 18 se debe a su lejanía del centro. Obsérvese que, para un modelo de regresión simple,

$$\sigma_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^2}$$

Como muestra el dibujo, $(x_{18} - \bar{x})^2$ es remarcadamente mayor que el resto de los $(x_j - \bar{x})^2$, ($j \neq 18$), ($\bar{x} = 14.38$).

Pese a que el caso 19 es un "outlier" no es excesivamente influyente según D_i . Su eliminación trasladaría $\hat{\beta}$ a la frontera de un elipsoide de confianza al 20%. Sin embargo, la eliminación del caso 18 efectuaría una traslación de $\hat{\beta}$ a la frontera de un elipsoide de confianza al 48%. El caso 2, como los demás casos, no presenta ninguna particularidad de importancia según D_i .

Los resultados obtenidos con D_i en cuanto a la ordenación de casos respecto a su influencia quedan confirmados por el estadístico de Andrews-Pregibon en el sentido de que R_{18} y R_{19} son los dos valores más pequeños de R_i . En la filosofía inicial de Andrews-Pregibon, el caso 18 es un "outlier" que importa y el caso 19 un "outlier" que no.

4. MULTICOLINEALIDAD EN REGRESION LINEAL

4.1. Definición

Los términos multicolinealidad, colinealidad y mala condición ("ill-conditioning") denotan una situación en la que las va-

riables regresoras del modelo mantienen una dependencia lineal aproximada. Los dos primeros son sinónimos y se usan indistintamente. El tercero es un concepto tomado del análisis numérico y trasladado al problema de mínimos cuadrados en el sentido de la mala condición del sistema de ecuaciones normales $X'X\beta = X'Y$ cuando las variables del modelo son colineales. Como señalan BKW (1980), la multicolinealidad no es propiamente un problema estadístico porque sólo depende de los datos en X pero es susceptible de disminuir la eficacia y fiabilidad de determinados aspectos del análisis mínimo-cuadrático. Hay que observar que, por su propia naturaleza, la colinealidad es un problema difícil de definir porque, en ocasiones, depende de la capacidad de cálculo disponible, (Weisberg, 1985). Se suele hablar de datos colineales o del grado de colinealidad presente en un conjunto de datos.

La multicolinealidad es un tema clásico de la investigación en modelos lineales y la bibliografía es muy extensa. Silvey (1969) discute los efectos de la colinealidad en la estimación de funciones paramétricas $X'\beta$. Hoerl y Kennard (1970) motivan la introducción de la regresión "ridge" mostrando la tendencia del estimador $\hat{\beta}$ hacia longitudes altas y signos incorrectos en presencia de multicolinealidad. Marquardt (1970) introduce el concepto de inflación de varianza. Marquardt y Snee (1975) recomiendan centrar y tipificar las variables regresoras para eliminar la mala condición no esencial ("non-essential ill-conditioning"). Otros trabajos interesantes se deben a Gunst - Mason (1977), Willan y Watts (1978) y Bradley-Srivastava (1979). Posteriores son los libros de Mason-Gunst (1980) y BKW (1980). El libro de Mason y Gunst incluye un capítulo dedicado a distintos métodos de estimación sesgada que pueden ser alternativos a los mínimos cuadrados cuando los datos muestran un alto grado de multicolinealidad. Entre ellos, la regresión "ridge", la regresión por componentes principales y la "latent root regression". Esta última se debe a Webster, Gunst y Mason (1974). El libro de BKW introduce varias técnicas nuevas y se manifiesta contrario a la práctica de centrar los datos. Trabajos más recientes son los de Mansfield y Helms (1982), Gunst (1983), Belsley (1984) y Gunst-Mason (1985).

La existencia de una relación lineal aproximada entre las

variables X del modelo puede representarse mediante un vector c , de longitud 1, tal que

$$Xc \approx 0 \quad (1.17)$$

Sea, en la notación usual,

$$X'X = CDC' \quad (1.18)$$

la descomposición espectral de la matriz $X'X$. De (1.17) se sigue que $c'X'Xc \approx 0$ lo que implica la existencia de autovalores pequeños de la matriz $X'X$ por ser (Rao, (1973))

$$\inf_{c'c=1} c'X'Xc = \lambda_{\min}[X'X] = \lambda_1 \quad (1.19)$$

Por consiguiente, la matriz de varianzas-covarianzas del estimador $\hat{\beta}$, $V[\hat{\beta}] = \sigma^2(X'X)^{-1}$ tiene elementos diagonales grandes al ser

$$\text{tr}[V[\hat{\beta}]] = \sigma^2 \sum_{j=1}^m \lambda_j^{-1}$$

Este razonamiento demuestra que el vector de coeficientes mínimo-cuadráticos es estadísticamente inestable ($\text{var}[\hat{\beta}_j]$ altas). Por otro lado, si $u = C'X'Y$, donde C es la matriz ortogonal $m \times m$ de (1.16), se puede poner

$$\hat{\beta} = \sum_{j=1}^m \frac{1}{\lambda_j} Y_j u_j \quad (1.20)$$

siendo $C = [Y_1, \dots, Y_m]$. Si $\lambda_1 \approx 0$, la fórmula (1.20) puede ser muy sensible ante pequeñas perturbaciones en X y en Y lo que demuestra que $\hat{\beta}$ está, también, afectado de problemas de inestabilidad numérica.

Un punto importante del trabajo de Gunst (1983) es la introducción de la primera definición formal de multicolinealidad. En analogía con el concepto de dependencia lineal de un conjunto de vectores $\{x_1, \dots, x_m\}$ de \mathbb{R}^n , Gunst propone la siguiente definición formal de multicolinealidad entre las columnas de la matriz X .

DEFINICION

Supóngase el modelo (1.1). Si para una cantidad $\eta > 0$ (fijada) existe $c \in \mathbb{R}^m$ no nulo tal que $\|Xc\| \leq \eta \|c\|$, entonces se dice que existe una colinealidad entre las columnas de la matriz X .

Esta definición, aunque útil, no es completamente satisfactoria. Esto subraya, una vez más, el carácter complejo de la definición exacta del problema de multicolinealidad en regresión lineal.

4.2. Causas y efectos

La multicolinealidad puede aparecer por distintas razones. En general, se distinguen cinco causas principales: especificación del modelo, multicolinealidades inherentes a la población, muestreo defectuoso, multicolinealidades provocadas por "outliers" o filas extremas de la matriz X y, por último, modelos sobredefinidos. Una vez que se ha decidido que la colinealidad procede de una de estas cinco causas, las soluciones o acciones que deben tomarse no son rutinarias y es difícil dar una norma de carácter general. Una discusión detallada, puede verse en Gunst (1983).

En cuanto a los efectos, estos son múltiples y variados, aunque, como ya se ha indicado anteriormente, pueden resumirse de modo global en problemas de inestabilidad numérica y estadística del estimador $\hat{\beta}$. Siguiendo de nuevo a Gunst (1983) las multicolinealidades son consideradas perjudiciales cuando: i) los valores de los estimadores y estadísticos muestrales están afectados más por las multicolinealidades que por el verdadero valor de los parámetros; ii) las varianzas de los estimadores y la longitud de los intervalos de confianza son excesivamente grandes o (iii) las predicciones son muy imprecisas cuando el punto en el que se predice no sigue la estructura de multicolinealidad de los datos. Las multicolinealidades no necesariamente producen estos problemas pero hay una fuerte tendencia de aparición de los mismos cuando las variables regresoras son colineales.

Si Y_i es un autovector normalizado correspondiente al autovalor λ_i y $\lambda_i \approx 0$, se tiene $X'X Y_i = \lambda_i Y_i \Rightarrow \|X Y_i\|^2 = \lambda_i \approx 0$ de donde

$$X Y_i \approx 0 \quad (1.21)$$

La relación (1.21) es una multicolinealidad descrita por las coordenadas altas de Y_i . En virtud de la descomposición espectral de (1.18) se tiene

$$\text{var}[\hat{\beta}_k] = \sigma^2 \sum_{j=1}^m \frac{Y_{kj}^2}{\lambda_j} \quad (1.22)$$

$$y \quad \text{cov} [\hat{\beta}_k, \hat{\beta}_l] = \sigma^2 \sum_{j=1}^m \frac{\gamma_{kj} \gamma_{lj}}{\lambda_j} \quad (1.23)$$

Por tanto, si las variables x_k y x_l son colineales, los sumandos dominantes en (1.22) y (1.23) son, respectivamente,

$$\begin{aligned} \text{var} [\hat{\beta}_k] &\approx \sigma^2 \frac{\gamma_{kj}^2}{\lambda_j} \\ \text{cov} [\hat{\beta}_k, \hat{\beta}_l] &\approx \sigma^2 \frac{\gamma_{kj} \gamma_{lj}}{\lambda_j} \end{aligned}$$

lo que permite concluir que, en tal caso, $\text{var}[\hat{\beta}_k]$ y $\text{var}[\hat{\beta}_l]$ son grandes así como $|\text{cov}[\hat{\beta}_k, \hat{\beta}_l]|$.

Cuando sólo hay un autovalor pequeño, la relación (1.20) implica que

$$\hat{\beta}_j \approx \frac{u_1}{\lambda_1} \gamma_{j1}$$

y, por tanto, si γ_{j1} es grande se tienen: i) estimadores grandes en magnitud salvo que $u_1 \approx 0$ y ii) signos determinados más por la multicolinealidad que por la relación de los regresores con la respuesta.

Considérese la variable aleatoria $L_1^2 = \|\hat{\beta} - \beta\|^2 = (\hat{\beta} - \beta)'(\hat{\beta} - \beta)$. El error cuadrático medio (E.C.M.) es la distancia esperada del estimador al parámetro,

$$\text{E.C.M.}(\hat{\beta}, \beta) = E[L_1^2] = \text{tr}[\sigma^2 (X'X)^{-1}] = \sigma^2 \sum_{j=1}^m \frac{1}{\lambda_j}$$

Si algún λ_j es pequeño, el E.C.M. de $\hat{\beta}$ en la estimación de β es grande. Por otro lado, puesto que $E[\hat{\beta}] = \beta$ y $V[\hat{\beta}] = \sigma^2 (X'X)^{-1}$, entonces

$$E[\|\hat{\beta}\|^2] = E[\hat{\beta}'\hat{\beta}] = \beta'\beta + \text{tr}[\sigma^2 (X'X)^{-1}] = \|\beta\|^2 + \sigma^2 \sum_{j=1}^m \frac{1}{\lambda_j}$$

Por tanto, en presencia de multicolinealidad, $\hat{\beta}$ es, en media, mucho más largo que β y $\|\hat{\beta}\|^2$ tiene mucho sesgo para estimar $\|\beta\|^2$ aún siendo $\hat{\beta}$ insesgado para β (Hoerl y Kennard, 1970).

Silvey (1969) estudia el problema de la estimación de combinaciones lineales $\lambda'\beta$. Si $X'X = CDC'$, $\lambda = C.a$, unitario, entonces

$$\text{var} [\lambda'\hat{\beta}] = \sigma^2 \lambda' (X'X)^{-1} \lambda = \sigma^2 a' D^{-1} a = \sigma^2 \sum_{j=1}^m \frac{a_j^2}{\lambda_j} \quad (1.24)$$

Silvey concluye que se pueden tener estimaciones relativamente precisas en las direcciones de los autovectores de $X'X$ asociados a autovalores grandes. Las estimaciones más imprecisas corresponden a los autovalores pequeños. Por convexidad en (1.24) se tiene:

$$\frac{\sigma^2}{\lambda_m} \leq \text{var} [\lambda'\hat{\beta}] \leq \frac{\sigma^2}{\lambda_1}$$

Los extremos de esta desigualdad se alcanzan con $\lambda = \gamma_m$ y $\lambda = \gamma_1$ respectivamente. La cota superior puede carecer de significado si λ_1 es excesivamente pequeño.

Willan y Watts (1978) desarrollan una medida para evaluar el efecto negativo de la multicolinealidad en los procedimientos de inferencia mínimo-cuadráticos. El desarrollo exige tipificar los datos $[Z]$ y comparar la longitud del intervalo de confianza de coeficiente $1-\alpha$ para α_j ($j=1, \dots, p$) (tomando $\alpha = (\alpha_1, \dots, \alpha_p)'$ en el modelo $E[Y] = I_n \alpha_0 + Z\alpha$) obtenido con Z con la longitud del mismo intervalo pero calculado con una matriz Z_0 , relativa a un hipotético o ficticio "diseño de referencia ortogonal" (Por tanto $Z_0'Z_0 = I_p$). Si $\hat{\sigma}^2$ es común para Z y Z_0 y, además $(Z'Z)^{-1} = (Z_{jj}^{-1})$, ($j=1, \dots, p$), las longitudes respectivas son

$$L = Z \hat{\sigma} t_{n,m, \frac{\alpha}{2}} \cdot (Z_{jj}^{-1})^{-1/2}$$

y

$$L_0 = Z \hat{\sigma} t_{n,m, \frac{\alpha}{2}}$$

Una posible medida de la pérdida de precisión en la estimación de α_j causada por la multicolinealidad es la razón

$$\frac{L_0}{L} = (Z_{jj}^{-1})^{-1/2} = (1 - R_j^2)^{1/2}$$

siendo R_j^2 el coeficiente de correlación múltiple entre z_j (la j -ésima columna de Z) y las demás z_k ($k \neq j$). Cuanto mayor es R_j^2 , mayor es la asociación lineal entre z_j y los z_k ($k \neq j$), mayor es L y menor la razón L_0/L . Implícito en este desarrollo está el hecho de que las dependencias lineales entre las columnas de Z tienden a provocar la aparición de elementos diagonales altos en $(Z'Z)^{-1}$

Otro efecto conocido de la multicolinealidad se manifiesta en los estadísticos t_j (o estadísticos F_j parciales) para contrastar $H_0: \alpha_j = 0$ frente a $H_1: \alpha_j \neq 0$. De acuerdo con la hipótesis lineal general

$$t_j = \frac{\hat{\alpha}_j}{\hat{\sigma} (\mathbf{z}^j \mathbf{z}^j)^{1/2}} = (\mathbf{z}^j \mathbf{z}^j)^{-1/2} \frac{\hat{\alpha}_j}{\hat{\sigma}}, \quad j = 1, \dots, p$$

Es frecuente observar que los estadísticos t_j asociados con variables colineales son pequeños debido a la inestabilidad provocada por los valores altos de $\mathbf{z}^j \mathbf{z}^j$. Por otro lado, la inestabilidad numérica puede hacer que $\hat{\alpha}_j$ sea grande, causando así un alto valor de t_j . Por consiguiente, la multicolinealidad origina que las variables involucradas en la misma puedan tener valores altos o bajos de t_j , independientemente de los valores auténticos de los parámetros. Esta ambigüedad en los valores de t_j no aparece al considerar la función de potencia de estos contrastes que es función monótona del parámetro de no centralidad de la distribución

$$\delta_j = (\mathbf{z}^j \mathbf{z}^j)^{-1/2} \frac{\alpha_j}{\sigma} \quad (j = 1, \dots, p)$$

Para valores de $\mathbf{z}^j \mathbf{z}^j$ altos, δ_j es bajo y con él la potencia del test. De aquí se concluye que la multicolinealidad también puede afectar a diversos métodos de selección de variables.

Los efectos en la predicción pueden ilustrarse como sigue. Si \hat{Y}_0 es la predicción correspondiente al vector x_0 , C es la matriz de la descomposición espectral y $x_0 = Cb$, la varianza de la predicción es

$$\text{var}[\hat{Y}_0] = \sigma^2 \left[\frac{1}{n} + x_0' (X'X)^{-1} x_0 \right] = \sigma^2 \left[\frac{1}{n} + \sum_{j=1}^m \frac{b_j^2}{\lambda_j} \right]$$

La predicción puede ser imprecisa ($\text{var}[\hat{Y}_0]$ alta) si $x_0 \approx x_j$, autovector asociado a un autovalor pequeño de $X'X$. Por otro lado, si x_0 tiene la misma estructura de multicolinealidad de los datos, x_0 está en la dirección de autovectores asociados a autovalores grandes de $X'X$ y la predicción tiende a ser razonable.

4.3. Diagnósticos de multicolinealidad

Los diagnósticos de multicolinealidad son indicadores o medidas del grado de condición de los datos. En la mayoría de las situaciones, los métodos de diagnóstico exigen una transformación de los datos encaminada a la consecución de la uniformidad de escalas en las columnas de la matriz de diseño. Usualmente, los datos se someten a una centralización y posterior tipificación. La centralización se utiliza para evitar las multicolinealidades debidas a distintos orígenes de las variables y la tipificación para eliminar los problemas que causan los diferentes escalas de los predictores. BKW (1980) se manifiestan contrarios a la práctica de la centralización y exigen tan solo la igualdad de escalas. Insistiendo en esto, Belsley (1984) argumenta que la centralización produce diagnósticos falsos de condición y enmascara el papel del término independiente en las dependencias lineales aproximadas. Por lo que se refiere a este trabajo, se seguirá siempre el criterio de centrar y tipificar los datos.

4.3.1. Coeficientes de correlación

Supóngase $Z'Z = (r_{jk})$, ($j, k = 1, \dots, p$). Los coeficientes de correlación r_{jk} ($j \neq k$) son indicadores clásicos para investigar la asociación entre pares de columnas de Z . Cuando $|r_{jk}| \rightarrow 1$ los vectores z_k y z_j tienden a la dependencia lineal perfecta. El inconveniente evidente de estas medidas es que no sirven para investigar dependencias lineales entre más de dos variables. Además, existen ejemplos de datos mal condicionados que no manifiestan su anomalía a través de los r_{jk} .

Otra medida relacionada son los coeficientes de correlación parcial. Si $(Z'Z)^{-1} = (r^{jk})$, el coeficiente de correlación parcial entre las variables j y k es

$$r_{jk} = - \frac{r_{jk}}{(r^{jj})^{1/2} (r^{kk})^{1/2}} \quad \left(\begin{array}{l} j \neq k \\ j, k = 1, \dots, p \end{array} \right)$$

Este diagnóstico constituye el fundamento de una técnica debida a Farrar y Glauber (1967) y, en esencia, refleja el grado de asociación lineal entre z_j y z_k una vez eliminado el efecto o de

pendencia de las demás columnas $(z_l, l \neq j, k)$.

Para analizar el grado de asociación lineal entre una variable z_j y las demás $z_k (k \neq j)$ puede utilizarse el coeficiente de correlación múltiple R_j^2 que no es más que el coeficiente de determinación de la regresión de z_j sobre los demás $z_k (k \neq j)$. Valores de $R_j^2 \geq 0.9$ son indicadores de problemas. Se tiene que

$$r_{jj} = (1 - R_j^2)^{-1} \quad (j = 1, 2, \dots, p)$$

4.3.2. Factores de inflación de la varianza

Marquardt (1970) introduce el concepto de factor de inflación de la varianza (V.I.F. = variance inflation factor) utilizando la representación

$$\text{var}[\hat{\alpha}_j] = \sigma^2 \text{VIF}_j = \sigma^2 r_{jj} = \sigma^2 \frac{1}{1 - R_j^2} \quad j = 1, \dots, p$$

Los $\text{VIF}_j (j = 1, \dots, p)$ son, pues, los elementos diagonales de la inversa $(Z'Z)^{-1}$. En ausencia de correlación lineal entre la variable z_j y los demás $z_k (k \neq j)$, $R_j^2 = 0$ y $\text{var}[\hat{\alpha}_j] = \sigma^2$. Por consiguiente, VIF_j puede interpretarse como el aumento en la varianza del estimador de α_j debido a la correlación entre z_j y los demás $z_k (k \neq j)$. Es fácil probar, usando la desigualdad de Cauchy-Schwartz, que

$$|r_{jk}| \leq [\text{VIF}_j \cdot \text{VIF}_k]^{1/2} \leq \max_j \text{VIF}_j$$

Por consiguiente, valores grandes de $\max_j \text{VIF}_j$ son indicadores de potencial inestabilidad numérica y estadística. Es usual admitir que valores de $\text{VIF}_j \geq 10$ son indicadores de la intervención de z_j en una dependencia lineal aproximada. En ocasiones, se usa $\text{VIF} = \max_j \text{VIF}_j$.

Un inconveniente de los $\text{VIF}_j (j = 1, 2, \dots, p)$ es que no ilustran la forma de la multicolinealidad al no ser capaces de detectar las variables involucradas en la misma.

4.3.3. Módulo de Gunst

La definición de multicolinealidad de Gunst puede usarse como fundamento de un método de diagnóstico. En efecto, la definición de Gunst es equivalente a afirmar que existe $c \in \mathbb{R}^m$,

$c'c = 1$, tal que $\|Xc\| \leq \eta$. A su vez, esto es equivalente a la condición $\inf_{c'c=1} \|Xc\| \leq \eta$. En efecto, si existe c unitario tal que $\|Xc\| \leq \eta$, entonces es inmediato que $\inf_{c'c=1} \|Xc\| \leq \eta$. Por otro lado, de la condición (1.19) es

$$\lambda_1 = \|X\gamma_1\|^2 = \inf_{c'c=1} \|Xc\|^2 = \inf_{c'c=1} c'X'Xc$$

de donde si $\inf_{c'c=1} \|Xc\|^2 \leq \eta$ la definición de Gunst se cumple tomando $c = \gamma_1$. De este razonamiento se deduce que existe multicolinealidad entre las columnas de la matriz X cuando $\lambda_1^{1/2} \leq \eta$ y recíprocamente. Por conveniencia, $\lambda_1^{1/2}$ será llamado, en ocasiones, modelo de multicolinealidad de Gunst. Implícito en este desarrollo, está el hecho de que si $\lambda_1^{1/2} \leq \eta$, la multicolinealidad asociada viene descrita por las coordenadas grandes del vector γ_1 .

El método anterior no es del todo satisfactorio. En primer lugar, la magnitud de η no está sujeta a una regla general que indique por debajo de qué punto crítico el módulo de multicolinealidad de Gunst es indicativo de problemas de degeneración en la matriz X . Por otro lado, la definición de Gunst, identificaría problemas de multicolinealidad en las matrices

$$X'X = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} \quad X'X = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \quad (1.25)$$

donde $0 < \alpha < \beta$ y $\alpha \rightarrow 0$. Esta situación es contradictoria porque las matrices (1.25) están perfectamente condicionadas al ser las columnas de X ortogonales. Esta dificultad se resuelve inmediatamente apelando a la dependencia de la definición de Gunst de las unidades de medida. Como apunta el propio Gunst (1984), si las columnas de la matriz X se tipifican para obtener longitudes unitarias, las matrices de productos cruzados (1.25) coinciden con la identidad de orden 2.

La definición de Gunst puede modificarse fácilmente para aplicarse en el contexto más usual de datos centrados y tipificados [matriz Z].

4.3.4. Autosistema de la matriz de correlación

Las medidas más eficaces y directas de multicolinealidad son los autovalores y autovectores de $Z'Z$. Si $\lambda_j \approx 0$, entonces $Z'Z y_j = \lambda_j y_j$ de donde $\|Z y_j\|^2 = \lambda_j \approx 0$ y $Z y_j \approx 0$. Por tanto, los autovalores pequeños detectan multicolinealidades que están descritas por las coordenadas grandes del correspondiente autovector. En ausencia de otro tipo de consideraciones, un punto de corte usual es declarar colineales aquellos autovalores $\lambda_j \leq 0.10$.

4.3.5. Diagnósticos de BKW

La idea de Belsley, Kuh y Welsch (1980) es analizar el conjunto de índices de condición

$$\eta_i = \left[\frac{\lambda_{\max}}{\lambda_i} \right]^{1/2}, \quad i=1, \dots, p$$

Valores altos de los η_i son indicadores de fuerte multicolinealidad.

El índice η_i es el número de condición de la matriz Z y se denotará por $k[Z]$. Puede probarse que $k[Z]$ es una medida de la sensibilidad potencial de la solución del sistema de ecuaciones $Z'Z\alpha = Z'Y$ ante perturbaciones o cambios de los datos $Z'Z$ y $Z'Y$. (Stewart (1973) y Berk (1977)). Para el número de condición un punto de corte razonable es $k[Z] > 30$.

Para analizar las variables que intervienen en una dependencia lineal aproximada, los índices de condición se complementan con las denominadas proporciones de descomposición de la varianza.

Si $Z'Z = CDC'$, como $V[\hat{\alpha}] = \sigma^2(Z'Z)^{-1}$ es

$$V[\hat{\alpha}] = \sigma^2(Z'Z)^{-1} = \sigma^2 C D^{-1} C' = \sigma^2 \sum_{j=1}^p \frac{1}{\lambda_j} y_j y_j' \Rightarrow$$

$$\text{var}[\hat{\alpha}_k] = \sigma^2 \sum_{j=1}^p \frac{y_{kj}^2}{\lambda_j} \quad (k=1, \dots, p)$$

Las proporciones de descomposición de la varianza son

$$\pi_{jk} = \frac{y_{kj}^2 / \lambda_j}{\sum_{l=1}^p (y_{kl}^2 / \lambda_l)} \quad \begin{matrix} j = 1, \dots, p \\ k = 1, \dots, p \end{matrix}$$

La matriz $\Pi = (\pi_{jk})$ se examina para averiguar la intervención de y_{kj}^2 / λ_j en la k -ésima varianza. Obsérvese que para λ_j

fijo, Π_{jk} es grande para coordenadas grandes $|Y_{kj}|$ de $Y_j = (Y_{1j}, \dots, Y_{kj}, \dots, Y_{nj})$.

4.4. Ejemplo

Chatterjee y Price (1977) estudian un modelo $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$, con $n = 11$ observaciones sobre la economía francesa. Las variables son

- Y = Importaciones
- x_1 = producción doméstica
- x_2 = formación de stocks
- x_3 = consumo doméstico

Las observaciones se tomaron entre los años 1949 y 1959, ambos inclusive. La tabla de análisis de la varianza resultante es

F. de V.	s. .	g.l.	C.M.	F
Regresión (corr.)	204,776	3	68,259	285,610
Error	1,673	7	0,239	
$R^2 = 0,992$				
Total (corr.)	106,449	10		

y los resultados usuales del análisis son

	$\hat{\beta}_j$	e. estándar($\hat{\beta}_j$)	t_j	$\hat{\alpha}_j$
x_1	-0.051	0.07	-0.731	-4.88
x_2	0.587	0.095	6.203	3.06
x_3	0.287	0.102	2.807	18.72

En este ejemplo, destaca inmediatamente el valor negativo del coeficiente de x_1 , en contradicción con conclusiones elementales de la teoría económica. También hay que resaltar la alta desviación típica estimada (0.07) de este coeficiente en relación con su valor estimado.

Los dos problemas señalados en el párrafo anterior son dos de los efectos típicos de la multicolinealidad porque indican inestabilidad numérica y estadística del coeficiente $\hat{\beta}_1$. Para confirmar si, efectivamente, es la multicolinealidad la responsable de estas anomalías, se estudian los correspondientes diagnósticos.

Matriz de correlación

	x_1	x_2	x_3
x_1	1	0.026	0.997
x_2	-	1	0.036
x_3	-	-	1

	$\lambda_1 = 0.00269$	$\lambda_2 = 0.99815$	$\lambda_3 = 1.9915$	VIF _i
x_1	0.707	-0.636	0.706	186.00
x_2	0.007	0.999	0.044	1.02
x_3	-0.707	-0.026	0.707	186.1

Todos los diagnósticos muestran una gran asociación lineal entre las variables x_1 y x_3 . Teniendo en cuenta que en presencia de un único autovalor pequeño se tiene

$$\hat{\alpha} \approx \frac{u_1}{\lambda_1} \gamma_1 \quad \text{donde} \quad u_1 = \gamma_1' Z' Y$$

y

$$\text{var}[\hat{\alpha}_1] \approx \sigma^2 \frac{\gamma_{11}^2}{\lambda_1}$$

puede, efectivamente, achacarse a la multicolinealidad la anomalía detectada en el coeficiente de x_1 . Obsérvese, finalmente, que el patrón de los signos de las coordenadas de $\hat{\alpha}$ es opuesto al patrón de los signos de γ_1 .

5. La matriz ampliada X^*

Uno de los elementos básicos sobre los que se desarrolla la investigación en este trabajo es la matriz

$$X^* = [X | Y] \quad (1.26)$$

la matriz de diseño del modelo (1.1) aumentada con la columna de observaciones Y . La matriz de (1.26) será la matriz aumentada o ampliada.

De modo esporádico, aunque con fines distintos a los de este trabajo, esta matriz aparece en diversos puntos de la teoría de modelos lineales. Webster, Gunst y Mason (1974), Gunst-Mason (1977) y Gunst-Mason (1980) desarrollan un método de estimación sesgada basado en el autosistema de la matriz $X^{*'}X^*$. En Seber (1977), pueden verse técnicas numéricas de cálculo relativas a la matriz $X^{*'}X^*$. En el trabajo ya citado de Andrews-Pregibon (1978) se desarrolla una medida de influencia basada en los cambios del volumen engendrado por $X^{*'}X^*$ al eliminar un caso del análisis. Recientemente, Gray-Ling (1985) hacen uso de los elementos de la matriz de proyección $V^* = X^*(X^{*'}X^*)^{-1}X^{*'}_1$ como instrumentos de detección de casos influyentes.

Uno de los objetivos de esta memoria es demostrar cómo la matriz X^* contiene información relevante: i) al ajuste del modelo (1.1), ii) a la multicolinealidad en X y iii) a la influencia de los casos. Así, la matriz X^* puede considerarse como instrumento central en un análisis de regresión por mínimos cuadrados.

Este apartado termina con la demostración de un resultado relativo a los elementos diagonales de la matriz de proyección V^* . Puesto que $X^* = [X | Y]$ puede aplicarse el proceso de ortogonalización (Cook-Weisberg (1982), M. del Río (1986)) para escribir

$$V^* = X^*(X^{*'}X^*)^{-1}X^{*'} = X(X'X)^{-1}X' + \frac{ee'}{SCE} = V + \frac{ee'}{SCE} \quad (1.27)$$

Según (1.27),

$$v_{ii}^* = v_{ii} + \frac{e_i^2}{SCE} = 1 - [1 - v_{ii}] \left[-\frac{r_i^2}{n-m} + 1 \right] \quad (1.28)$$

Por consiguiente, σ_{ii}^* es una función monótona creciente de las coordenadas del bloque básico de influencia (σ_{ii}, r_i^2) y puede considerarse como una medida de influencia. De hecho,

$$R_i = 1 - \sigma_{ii}^*$$

El resultado (1.28) será de utilidad en los Capítulos III y IV.

CAPITULO II

1. SUMARIO

Este capítulo es un estudio de tres tipos distintos de diagnósticos de multicolinealidad sobre la matriz aumentada y tipificada Z^* y su relación con los correspondientes diagnósticos de la matriz tipificada original Z .

La sección 2 contiene una relación de propiedades de la matriz Z^* y de la matriz de productos cruzados $Z^{*'}Z^*$.

En la sección 3 se analizan los factores de inflación de la varianza asociados a Z^* . El apartado 3.1 propone dos métodos de detección de multicolinealidad en Z con la información deducida de estos factores.

La sección 4 se dedica al estudio del autosistema de la matriz $Z^{*'}Z^*$. Se obtiene un importante teorema de separación en el apartado 4.1.1. que permite establecer desigualdades entre los índices de condición de $Z'Z$ y $Z^{*'}Z^*$. En el apartado 4.1.2. se deduce la distribución del menor de los autovalores de la matriz de productos cruzados ampliada en un caso sencillo y, por último, en el apartado 4.2. se demuestra una fórmula para el j -ésimo autovector normalizado de $Z^{*'}Z^*$. El apartado 4.3. contiene una relación entre el número de condición de $Z^{*'}Z^*$ y los elementos diagonales de su inversa.

La sección 5 se compone de un desarrollo destinado a la obtención de una desigualdad, y su posterior aplicación, relativa a los módulos de multicolinealidad de Gunst asociados a Z y Z^* .

Finalmente, la sección 6 es un ejemplo práctico destinado a ilustrar los resultados obtenidos.

2. FORMULAS BASICAS. INTRODUCCION

Si se considera el modelo transformado de (1.1)

$$Y = 1_n \alpha_0 + Z \alpha + \varepsilon \quad (2.1)$$

la matriz ampliada (centrada y tipificada) Z^* está definida por

$$Z^* = [Z | \gamma] \quad (2.2)$$

siendo γ el vector de observaciones centradas y tipificadas, es decir $\gamma = [Y - 1_n \bar{Y}] / S_Y$, donde $\bar{Y} = \sum Y_i / n$ y $S_Y^2 = \sum (Y_i - \bar{Y})^2$. La información deducida de la matriz Z^* de orden $n \times m$ ($m = p+1$) es el punto básico de este capítulo.

La matriz aumentada de productos cruzados $Z^{*'} Z^*$ presenta la siguiente estructura:

$$Z^{*'} Z^* = \begin{pmatrix} Z'Z & Z'\gamma \\ \gamma'Z & 1 \end{pmatrix} \quad (2.3)$$

Esta matriz, de orden $n \times m$, es la matriz de correlaciones entre todas las columnas de Z^* y contiene como submatriz principal $p \times p$ la matriz de correlaciones asociada al modelo (1.1).

Bajo la hipótesis usual de normalidad, la matriz Z^* es de rango máximo (c.s.). En efecto, basta considerar la cadena de equivalencias:

$$\gamma \in C[Z] \Leftrightarrow Y - 1_n \bar{Y} \in C[Z] \Leftrightarrow Y \in C[1_n | Z] \Leftrightarrow Y \in C[X]$$

que termina en un suceso de probabilidad cero si $n > m$. Por consiguiente, la matriz $Z^{*'} Z^*$ es definida positiva con probabilidad uno. Los autovalores de esta matriz se denotarán por

$$0 < \lambda_1^* \leq \lambda_2^* \leq \dots \leq \lambda_m^*$$

y por $\gamma_j^* = [\gamma_{1j}^*, \gamma_{2j}^*, \dots, \gamma_{nj}^*]'$ ($j=1, \dots, m$) los correspondientes autovectores normalizados.

Usando resultados conocidos (A.1), el determinante de (2.3) es

$$|Z^{*'} Z^*| = |Z'Z| [1 - \gamma'Z(Z'Z)^{-1}Z'\gamma] = |Z'Z| [1 - R^2] \quad (2.4)$$

La razón entre determinantes es, según (2.4), función monótona del coeficiente de determinación R^2 .

La inversa de (2.3) es

$$(Z^*Z^*)^{-1} = \begin{pmatrix} (Z'Z)^{-1} + \frac{\hat{\alpha}\hat{\alpha}'}{SCE} & -\frac{S_y}{SCE} \hat{\alpha} \\ -\frac{S_y}{SCE} \hat{\alpha}' & \frac{S_y^2}{SCE} = \frac{1}{1-R^2} \end{pmatrix} \quad (2.5)$$

donde $\hat{\alpha} = (Z'Z)^{-1} Z'Y$ es el estimador de α en el modelo (2.1). Conociendo entonces el valor de S_y , la matriz Z^*Z^* y su inversa proporcionan información sobre: i) el valor de SCE, ii) el valor de $\hat{\alpha}$, iii) el valor de R^2 , iv) el determinante $|Z'Z|$ y v) las correlaciones entre los regresores y entre éstos y la respuesta.

La diferencia clave entre la detección de multicolinealidades en Z y Z^* radica en la influencia que la respuesta Y ejerce en las multicolinealidades de la matriz ampliada. Sean por ejemplo, λ_j^* y γ_j^* dos elementos correspondientes del auto sistema de Z^*Z^* . Si $\lambda_j^* \approx 0$, entonces

$$\gamma_{1j}^* z_1 + \dots + \gamma_{pj}^* z_p + \gamma_{mj}^* Y \approx 0 \quad (2.6)$$

Siguiendo a Gunst y Mason (1980), cuando γ_{mj}^* no está muy cerca no a cero, la multicolinealidad (2.6) es predictiva porque se puede escribir

$$Y \approx -\frac{1}{\gamma_{mj}^*} [\gamma_{1j}^* z_1 + \dots + \gamma_{pj}^* z_p]$$

Cuando $\lambda_j^* \approx 0$, $\gamma_{mj}^* \approx 0$, la multicolinealidad es no predictiva y se refiere solo a las columnas de la matriz Z .

El párrafo anterior sugiere que un estudio de los diagnósticos de multicolinealidad en Z^* debe considerar como puntos fundamentales la detección de multicolinealidades en la matriz ampliada, el papel de la respuesta Y en los mismos y la relación de las dependencias lineales en Z^* con las dependencias en Z . Este programa básico se desarrolla en este capítulo en tres diagnósticos distintos: los factores VIF_j^* ($j = 1, \dots, m$), el autosistema de Z^*Z^* y el módulo de Gunst asociado a las columnas de Z^* .

3. ANALISIS DE LOS VIF_j^* ($j=1,2,\dots,m$).

En notación evidente, los VIF_j^* ($j=1,\dots,m$) son los elementos diagonales de la inversa de Z^*Z^* . Estos factores de inflación de la varianza constituyen un conjunto de diagnósticos que permite obtener los resultados más satisfactorios en relación con el problema considerado en este capítulo.

Según la inversa dada en (2.5), se sigue inmediatamente que

$$VIF_j^* = VIF_j + \frac{\hat{\alpha}_j^2}{SCE} \quad (j=1,2,\dots,p) \quad (2.7)$$

y

$$VIF_m^* = \frac{S_y^2}{SCE} = \frac{1}{1-R^2} \quad (2.8)$$

Según (2.8) un VIF_m^* alto es equivalente a una gran calidad del ajuste.

Mucho más interesante que (2.8) es el resultado de (2.7).

Teniendo en cuenta que el j -ésimo estadístico F parcial es $F_j = \hat{\alpha}_j^2 / (VIF_j) \hat{\sigma}^2$, las fórmulas (2.7) permiten escribir

$$VIF_j^* = VIF_j \left[1 + \frac{\hat{\alpha}_j^2}{(VIF_j) SCE} \right] = VIF_j \left[1 + \frac{1}{n-m} F_j \right] \quad j=1,\dots,p \quad (2.9)$$

En virtud de (2.9), el estadístico F_j actúa como un coeficiente de dilatación (F_j no tiene unidades) en el paso de VIF_j a VIF_j^* .

Finalmente, sea $(Z^*Z^*)^{-1} = (Z^{*jk})$ ($j,k=1,\dots,m$). Según (2.7) y (2.8) (y (2.5)) es

$$Z^{*jj} = VIF_j + \frac{(Z^{*jm})^2}{Z^{*mm}} \quad j=1,\dots,p \quad (2.10)$$

Por consiguiente, la inversa de la matriz Z^*Z^* permite calcular, además de la información citada en el apartado 2, i) los estadísticos $VIF_j^* = Z^{*jj}$, $j=1,\dots,p$ y $VIF_m^* = Z^{*mm}$, este último equivalente a R^2 , ii) los diagnósticos VIF_j ($j=1,\dots,p$) en virtud de la fórmula (2.10) y iii) los estadísticos F_j parciales. En efecto, de (2.7), (2.9) y (2.10) se deduce

$$F_j = (n-m) \frac{VIF_j^* - VIF_j}{VIF_j} = (n-m) \frac{\frac{(Z^{*j})^2}{Z^{*mm}}}{Z^{jj} - \frac{(Z^{*j})^2}{Z^{*mm}}} =$$

$$= (n-m) \frac{1}{\frac{Z^{jj} Z^{*mm}}{(Z^{*j})^2} - 1}$$

para $j=1, \dots, p$.

3.1. Detección de multicolinealidad en Z con los VIF_j^* ($j=1, 2, \dots, p$)

La representación (2.9) factoriza VIF_j^* en dos componentes. Una de ellas, VIF_j , es atribuible a la multicolinealidad en Z y la otra, $1 + \frac{1}{n-m} F_j$, a la intervención de Y en Z^* . Se deduce inmediatamente: i) si todos los VIF_j^* ($j=1, \dots, p$) son pequeños, los VIF_j ($j=1, \dots, p$) también lo serán y las matrices Z y Z^* no presentan problemas de dependencias lineales entre sus columnas; ii) Si VIF_j es grande, VIF_j^* es todavía mayor; iii) Un VIF_j^* puede ser grande debido bien a un VIF_j grande o bien debido a un gran valor de F_j .

La conclusión iii) sugiere que para detectar multicolinealidad en Z usando la información de Z^* basada en los VIF_j^* es necesario separar en (2.9) la parte de VIF_j y la parte de F_j . Obsérvese también que en iii) las dos causas citadas no son excluyentes porque la multicolinealidad en Z puede causar un F_j inusualmente grande. Para evitar ambigüedades en el análisis pueden darse dos posibles métodos.

3.1.1. Método directo

Un método simple y directo de identificación de multicolinealidad en Z con los VIF_j^* ($j=1, 2, \dots, p$) podría ser el siguiente:

Sea $k > 0$ un punto de corte para los VIF_j :

i) Si $VIF_j^* < k$ ($j=1, \dots, p$), entonces $\max_j VIF_j < k$

y no hay multicolinealidad en Z .

ii) Si para algún $VIF_j^* \gg k$, puesto que según (2.10) es

$$VIF_j = z^{*jj} - \frac{(z^{*jm})^2}{z^{*mm}} \quad (j=1, \dots, p)$$

entonces es $VIF_j \gg k$ si, y sólo si

$$\frac{(z^{*jm})^2}{z^{*mm}} \leq z^{*jj} - k \quad (j=1, \dots, p) \quad (2.11)$$

iii) En el caso de ser $VIF_j^* \gg k$ pero $\frac{(z^{*jm})^2}{z^{*mm}} > z^{*jj} - k$, es $VIF_j \gg k$ lo que indicaría que el alto valor de VIF_j^* se debe a una gran magnitud de F_j .

3.1.2. Distribución de los VIF_j^* ($j=1, 2, \dots, p$)

Bajo el modelo (1.1) con la hipótesis usual de normalidad, $\varepsilon \sim N_n(0, \sigma^2 I_n)$, la distribución de F_j es $F_{1, n-m}(\delta_j)$ donde

$$\delta_j = (VIF_j)^{-1} \frac{\alpha_j^2}{2\sigma^2} \quad j=1, \dots, p$$

La distribución de VIF_j^* se deduce inmediatamente de la relación (2.9). Manejar la distribución F no central es complicado y parece conveniente acudir a una aproximación basada en la F central. Se tiene (Patnaik (1949))

$$F_{q_1, q_2}(\delta) \approx \frac{q_1 + 2\delta}{q_1} F_{q^*, q_2} \quad (2.12)$$

siendo $q^* = (q_1 + 2\delta)^2 / (q_1 + 4\delta)$. Evidentemente, q^* puede tomar valores no enteros.

Sea la variable aleatoria $V^* = \max_{j=1, \dots, p} VIF_j^*$ y v^* su valor observado. Conocida la distribución de los VIF_j^* ($j=1, \dots, p$) puede diseñarse un método basado en la aproximación (2.12). Sustituyendo los parámetros δ_j por sus estimadores naturales

$$\hat{\delta}_j = (VIF_j)^{-1} \frac{\hat{\alpha}_j^2}{2\hat{\sigma}^2} \quad j=1, \dots, p$$

se tiene, en virtud de la desigualdad de Bonferroni,

$$\begin{aligned} p &= P_r \{V^* > v^*\} \leq \sum_{j=1}^p P_r \{VIF_j^* > v^*\} = \\ &= \sum_{j=1}^p P_r \{F_{1, n-m}[\delta_j] > (n-m)[w_j^* - 1]\} \approx \\ &\approx \sum_{j=1}^p P_r \{F_{1, n-m}[\hat{\delta}_j] > (n-m)[w_j^* - 1]\} \end{aligned} \quad (2.13)$$

donde $\omega_j^* = \sigma^* / \sqrt{1F_j^*}$ ($j=1, \dots, p$). Usando en (2.13) el resultado de (2.12) se sigue que, aproximadamente, p es menor o igual que

$$\sum_{j=1}^p P_r \left\{ F_{q^*j, n-m} > \frac{(n-m)[\omega_j^* - 1]}{1 + 2\hat{\delta}_j} \right\} \quad (2.14)$$

siendo $q^*j = (1 + 2\hat{\delta}_j)^2 / (1 + 4\hat{\delta}_j)$. Tablas de la distribución F central con grados de libertad no enteros pueden verse en Graybill (1976).

Si σ^* es un punto por debajo del cual la matriz Z^* se juzga bien condicionada entonces cuando la cota de (2.14) sea pequeña se tiene la confirmación de que, salvo en casos extremos, es innecesaria la investigación detallada de la multicolinealidad en Z .

Un inconveniente de este método, aparte de las aproximaciones citadas, es que la multicolinealidad en Z^* se calibra en términos globales mediante el nivel de significación p . Obsérvese que la mala condición en Z^* , en el sentido citado en la introducción, se define sin referencia a ningún tipo de distribución. No está clara la relación explícita entre un valor bajo de p y la ausencia de problemas de degeneración en la matriz aumentada.

4. AUTOSISTEMA DE LA MATRIZ DE PRODUCTOS CRUZADOS Z^*Z^* .

En lo que sigue, se utilizan las descomposiciones espectrales de $Z'Z$ y Z^*Z^*

$$Z'Z = CDC', \quad Z^*Z^* = C^*D^*C^{*'}.$$

donde en $C = [\gamma_1, \dots, \gamma_p]$, $D = \text{diag}(\lambda_1, \dots, \lambda_p)$, $C^* = [\gamma_1^*, \dots, \gamma_p^*, \gamma_m^*]$ y $D^* = \text{diag}(\lambda_1^*, \dots, \lambda_p^*, \lambda_m^*)$.

4.1. Autovalores

4.1.1. Teorema de separación. Consecuencias.

Uno de los principales resultados de este capítulo es la obtención de un teorema de separación entre los autovalores de $Z'Z$ y Z^*Z^* . La demostración utiliza técnicas de un resultado similar del libro de Wilkinson (1965).

TEOREMA 2.1.

Si los autovalores de $Z'Z$ son todos distintos entre sí, con probabilidad uno se cumple

$$0 < \lambda_1^* < \lambda_1 < \lambda_2^* < \lambda_2 < \dots < \lambda_p < \lambda_m^*$$

DEMOSTRACION:

La matriz $m \times m$

$$\tilde{C} = \begin{pmatrix} C & 0 \\ 0 & 1 \end{pmatrix}$$

es ortogonal por lo que los autovalores de Z^*Z^* y de $\tilde{C}'Z^*Z^*\tilde{C}$ coinciden. De manera explícita:

$$\begin{aligned} \tilde{C}'Z^*Z^*\tilde{C} &= \begin{pmatrix} C' & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Z'Z & Z'\gamma \\ \gamma'Z & 1 \end{pmatrix} \begin{pmatrix} C & 0 \\ 0 & 1 \end{pmatrix} = \\ &= \begin{pmatrix} D & u \\ u' & 1 \end{pmatrix} \end{aligned} \quad (2.15)$$

donde $u = C'\gamma$. El polinomio característico de la matriz (2.15) es

$$P[\lambda^*] = \begin{vmatrix} D - \lambda^* I_p & u \\ u & 1 - \lambda^* \end{vmatrix} \quad (2.16)$$

Es fácil ver que la distribución de u es una función de Student invertida multivariante (Raiffa-Schlaiffer 1961), de modo que $P[u_j \neq 0, j=1, \dots, p] = 1$, es decir, se puede afirmar que, con probabilidad 1, las coordenadas de u son todas distintas de cero. Por consiguiente, ningún autovalor de $Z'Z$ anula (2.16) porque $P[\lambda_j^*] = \pm u_j^2 \prod_{k \neq j} (\lambda_k - \lambda_j^*)$ es no nulo al ser los autovalores de $Z'Z$ distintos entre sí.

Las soluciones de la ecuación $P[\lambda^*] = 0$ coinciden, según el párrafo anterior, con las de

$$|D - \lambda_j^* I_p| [(1 - \lambda^*) - u' [D - \lambda_j^* I_p]^{-1} u] = 0$$

o, puesto que $|D - \lambda_j^* I_p| \neq 0$, con las soluciones de

$$\begin{aligned} 1[\lambda^*] &= \lambda^* + u' [D - \lambda_j^* I_p]^{-1} u = \\ &= \lambda^* + \sum_{j=1}^p \frac{u_j^2}{\lambda_j - \lambda^*} = 1 \end{aligned}$$

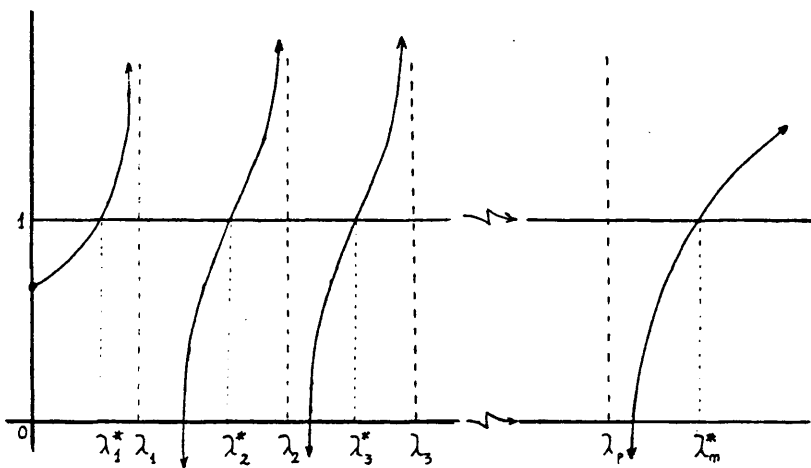


Fig 2.1

Dado que el gráfico de $f[\lambda^*]$ es de la forma indicada en la figura 2.1, se deduce inmediatamente la conclusión del teorema. #

El teorema anterior permite establecer con facilidad desigualdades entre los índices de condición de las matrices Z y Z^* . En efecto sean

$$\eta_k = \left[\frac{\lambda_{\max}}{\lambda_k} \right]^{1/2} \quad \text{y} \quad \eta_k^* = \left[\frac{\lambda_{\max}^*}{\lambda_k^*} \right]^{1/2}$$

$k = 1, \dots, p$ $k = 1, \dots, p, p+1$

los índices de condición de las matrices Z y Z^* , respectivamente. Sean, también,

$$\delta_k^* = \left[\frac{\lambda_p^*}{\lambda_{k+1}^*} \right]^{1/2}, \quad k = 1, \dots, p-1$$

Del teorema de separación es fácil ver que

$$\frac{\lambda_p^*}{\lambda_{k+1}^*} < \frac{\lambda_p}{\lambda_{k+1}} < \frac{\lambda_m^*}{\lambda_k^*}, \quad k = 1, \dots, p-1$$

y, por tanto, $\delta_{k+1}^* < \eta_k < \eta_k^*$, $k = 1, \dots, p-1$ (2.17). En particular, tomando $k=1$ en (2.17), se tiene la desigualdad correspondiente para el número de condición de la matriz Z

$$\delta_1^* < k[Z] < k[Z^*] \quad (2.18)$$

De (2.18) se deduce que si λ_1^* es pequeño en relación con λ_p^* la condición $k[Z]$ es mala. Por otro lado se obtiene también la evidente conclusión de que si $k[Z]$ es grande, $k[Z^*]$ es todavía mayor. Si λ_p^* y λ_m^* están próximos, las cotas (2.13) proporcionan un intervalo preciso para obtener información sobre $k[Z]$.

Pese a que el teorema de separación es un resultado interesante por sí mismo, la línea de consecuencias deducible del mismo conduce forzosamente a la obtención de desigualdades. Este procedimiento no es completamente satisfactorio porque, entre otras cosas, así es imposible delimitar con precisión el papel de la respuesta Y en los índices de condición de la matriz Z^* . Este tipo de dificultad será frecuente en el tratamiento de problemas relativos al autosistema de la matriz de productos cruzados.

4.12. Distribución de los autovalores

Webster, Gunst y Mason (1974) señalan que la distribución de los autovalores y autovectores de la matriz $Z^{*'}Z^*$ es difícil de obtener. Sin pretender haber resuelto el problema ni las cuestiones que éste lleva consigo: aproximaciones a la distribución, comportamiento asintótico, ..., etc., en este trabajo se aporta, por métodos simples de cambio de variable, una breve contribución al mismo. En esencia, se obtiene la expresión explícita de la densidad del menor de los autovalores de la matriz $X^{*'}X^*$, la matriz ampliada de productos cruzados del modelo (1.1), en un caso particular sencillo, pero ilustrativo de las dificultades del problema

Supóngase $Y = X\beta + \varepsilon$, donde X es un vector de longitud unidad y $\varepsilon \sim N_n(0; I_n)$. Los autovalores de $X^{*'}X^*$ son las raíces del polinomio

$$\begin{vmatrix} 1 - \lambda^* & \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n x_i y_i & SCT - \lambda^* \end{vmatrix}$$

o soluciones de la ecuación $SCT = \lambda^* + \frac{u^2}{1 - \lambda^*}$ donde $u = \sum_{i=1}^n x_i y_i$.

Pero $u^2 = SCR$ de modo que esta ecuación es equivalente a

$$SCE = \lambda^* + \frac{SCR}{1-\lambda^*} - SCR = \lambda^* \left[1 + \frac{SCR}{1-\lambda^*} \right] \quad (2.19)$$

Si $SCE = \sigma$ y $SCR = \omega$, la densidad conjunta $f(\sigma, \omega)$ es fácil de calcular. Usando (2.19), el cambio de variable

$$\begin{aligned} \sigma &= \lambda^* \left[1 + \frac{\omega}{1-\lambda^*} \right] \\ \omega &= [1-\lambda^*] t \end{aligned}$$

permite conocer la densidad conjunta de (λ^*, t) concentrada en el caso del menor de los autovalores en la región $0 < \lambda^* < 1$, $t > 0$ (ver tema de separación 2.1) que resulta ser

$$f[\lambda^*, t] = \left(\frac{1}{2}\right)^r \lambda^{*r-1} [(1-\lambda^*) + t] \cdot e^{-1/2(c + \lambda^* + t)} \quad (2.20)$$

$$\cdot \sum_{i=0}^{\infty} \frac{\Gamma(r) \left(\frac{c}{2}\right)^i}{i!} \cdot \frac{1}{2^{1+2i/2} \Gamma\left(\frac{1+2i}{2}\right)} (1+t)^{r-1} [(1-\lambda^*)t]^{1+i}$$

donde $c = \beta^2$ y $r = n-1/2$. La densidad del menor de los autovalores de $X^* X^*$ es $f[\lambda^*] = \int_0^\infty f[\lambda^*, t] dt$. En virtud de la estructura de (2.20) es útil la relación (Abramowitz-Stegun, 1964, pág. 505)

$$\Gamma(a) \cdot U(a, b, z) = \int_0^\infty e^{-zt} t^{a-1} (1+t)^{b-a-1} dt \quad (2.21)$$

en la que $a > 0$, $z > 0$ y $U(a, b, z)$ es la función hipergeométrica confluyente de Kummer. Usando (2.21) para resolver la integral se llega, tras las correspondientes operaciones, a la expresión explícita de $f[\lambda^*]$

$$\begin{aligned} f[\lambda^*] &= \left(\frac{1}{2}\right)^r e^{-c/2} \lambda^{*r-1} e^{-1/2 \lambda^*} \cdot \left[\frac{(1-\lambda^*)^{1/2}}{2^{1/2}} U\left(\frac{1}{2}, r + \frac{1}{2}, \frac{1}{2}\right) \right. \\ &\quad \left. + \sum_{i=0}^{\infty} \frac{U\left(\frac{1}{2} + \frac{3}{2}i, r + i + \frac{3}{2}, \frac{1}{2}\right)}{i! 2^{3/2+i}} \left(\frac{c}{2}\right)^i (1-\lambda^*)^{i-1/2} (2i+1) + \frac{c}{i+1} (1-\lambda^*)^i \right] \end{aligned} \quad (2.22)$$

si $0 < \lambda^* < 1$ y $c = \beta^2$.

La complejidad de (2.22) no permite dar respuesta a interrogantes tan elementales como son el conocer $E[\lambda^*]$ o $\text{var}[\lambda^*]$ o, por ejemplo, tener una idea aproximada del gráfico de la densidad $f[\lambda^*]$. Cabe sospechar que al aumentar la complejidad del modelo de partida, la correspondiente densidad $f[\lambda^*]$, caso de que pueda llegarse a una expresión compacta, será más intratable que (2.22).

4.2. Autovectores

Puede darse una relación explícita entre el j -ésimo autovector de $Z^{*'}Z^*$ y el autosistema de $Z'Z$.

Como paso previo, considérense los vectores $m \times 1$

$$\bar{y}_j = \begin{pmatrix} -[D - \lambda_j^* I_p]^{-1} u \\ 1 \end{pmatrix} \quad j = 1, \dots, p, m$$

Se trata de demostrar que \bar{y}_j es un autovector de la matriz (2.15) asociado al autovalor λ_j^* . En efecto se tiene

$$\begin{pmatrix} D & u \\ u' & 1 \end{pmatrix} \bar{y}_j = \begin{pmatrix} u - D[D - \lambda_j^* I_p]^{-1} u \\ 1 - u'[D - \lambda_j^* I_p]^{-1} u \end{pmatrix} \quad (2.23)$$

Ahora bien, como

$$\begin{aligned} u - D[D - \lambda_j^* I_p]^{-1} u &= u - [(D - \lambda_j^* I_p) + \lambda_j^* I_p] [D - \lambda_j^* I_p]^{-1} u = \\ &= -\lambda_j^* [D - \lambda_j^* I_p]^{-1} u \end{aligned}$$

y según el teorema de separación es

$$1 = \lambda_j^* + u'[D - \lambda_j^* I_p]^{-1} u$$

el segundo miembro de (2.23) es $\lambda_j^* \bar{y}_j$.

En virtud del párrafo anterior, los autovectores de $Z^{*'}Z^*$ son proporcionales a las columnas de la matriz $\tilde{C}[\bar{y}_1, \bar{y}_2, \dots, \bar{y}_p, \bar{y}_m]$. En concreto, la columna j -ésima de esta matriz es

$$\begin{pmatrix} C & 0 \\ 0 & 1 \end{pmatrix} \bar{y}_j = \begin{pmatrix} -C[D - \lambda_j^* I_p]^{-1} u \\ 1 \end{pmatrix} \quad (2.24)$$

$j=1, \dots, m$.

Teniendo en cuenta que \tilde{C} es una matriz ortogonal, la longitud del vector (2.24) es $1 + u'[D - \lambda_j^* I_p]^{-2} u$ y, de aquí, un conjunto de autovectores normalizados para la matriz $Z^{*'}Z^*$ es $[\gamma_1^*, \dots, \gamma_p^*, \gamma_m^*]$ donde

$$\gamma_j^* = \frac{1}{[1 + u'[D - \lambda_j^* I_p]^{-2} u]^{1/2}} \begin{pmatrix} -C[D - \lambda_j^* I_p]^{-1} u \\ 1 \end{pmatrix} \quad (2.25)$$

$j = 1, 2, \dots, p, m$

La expresión (2.25) combina información de la matriz $Z'Z$ [C y D] y de la matriz $Z^*Z^*(\lambda_j^*, u)$ y no puede, por tanto, utilizarse para determinar los Y_j^* en función de los Y_j . Debido a su complejidad, no resulta de utilidad tampoco para analizar las proporciones de descomposición de la varianza de Z^* ni el papel de la respuesta Y en las coordenadas de Y_j^* . No obstante, la fórmula (2.25) será de utilidad en otro contexto.

4.3. Desigualdades de Berk

Berk (1977) establece un resultado que es de utilidad en el estudio de los diagnósticos de multicolinealidad en Z^* .

TEOREMA 2.2. Berk (1977)

Si C es una matriz de correlación $m \times m$ no singular y $V_1 \leq V_2 \leq \dots \leq V_m$ son los elementos diagonales de la inversa, entonces

$$V_m \leq k^2 \leq m (V_1 + \dots + V_p + V_m)$$

donde k es el número de condición de C .

DEMOSTRACION:

Como $V_1 + \dots + V_p + V_m = \text{tr}[C^{-1}] = \frac{1}{\lambda_1[C]} + \dots + \frac{1}{\lambda_m[C]} \gg \frac{1}{\lambda_{\min}[C]}$ y por ser $\sum_{j=1}^m \lambda_j[C] = m$ es $\lambda_{\max}[C] \leq m$, entonces $k^2 \leq m (V_1 + \dots + V_p + V_m)$. Para llegar a la cota inferior, hay que tener en cuenta que al ser $\sum_{j=1}^m \lambda_j[C] = m$ es $\lambda_{\max}[C] \geq 1$ y puesto que $V_m \leq \max_{a' a = 1} a' C^{-1} a = 1/\lambda_{\min}[C]$ se tiene, finalmente,

$$V_m \leq \frac{\lambda_{\max}[C]}{\lambda_{\min}[C]} = k^2$$

Berk comenta que la cota inferior se alcanza cuando C es la matriz identidad mientras que la cota superior es precisa cuando C tiene un autovalor dominante, los demás pequeños y uno de ellos sensiblemente más pequeño que el resto.

Este teorema permite relacionar directamente los VIF_j^* y el cuadrado de la condición $k^2[Z^*]$. No hay más que recordar que Z^*Z^* es una matriz de correlación definida positiva con probabilidad uno, de donde

$$\max_j \text{VIF}_j^* \leq k^2 [Z^*] \leq m \left[\sum_{j=1}^m \text{VIF}_j^* \right]$$

En particular, $[1-R^2]^{-1} \leq k^2 [Z^*]$ lo que prueba que ajustes buenos provocan mala condición de la matriz Z^* .

5. MODULO DE GUNST

Entre las columnas de la matriz Z^* existe una multicolinealidad cuando para un cierto η^* fijo, existe c^* unitario tal que $\|Z^* c^*\| \leq \eta^*$. Es sabido que esto es equivalente a la condición $\lambda_1^{1/2} \leq \eta^*$. $\lambda_1^{1/2}$ es el módulo de multicolinealidad de Gunst de la matriz Z^* .

5.1. Relación entre los módulos de Z y Z^*

Evidentemente, si se satisface $\|Zc\| \leq \eta$ donde $c \in \mathbb{R}^p$, $c'c=1$ y $\eta > 0$ es una cantidad fijada, se satisface inmediatamente $\|Z^* c^*\| \leq \eta^*$ con $c^* = (c', 0)'$ y $\eta^* = \eta$. Dicho de otra forma, la existencia de una multicolinealidad en Z de módulo menor o igual que η describe, automáticamente, una multicolinealidad entre las columnas de Z^* de módulo menor o igual que η . No obstante, el problema que se pretende resolver es precisamente el contrario: establecer condiciones que permitan garantizar cuando $\lambda_1^{1/2}$ pequeño implica $\lambda_1^{1/2}$ pequeño.

Si $\gamma_{\min}^* = (\gamma_1^*, \gamma_m^*)'$, donde γ_1^* es $p \times 1$, es el autovector normalizado correspondiente a λ_1^* se tiene $\|Z^* \gamma_{\min}^*\|^2 = \lambda_1^*$. Considérese la combinación lineal

$$Z^* \gamma_{\min}^* = [Z | Y] \begin{pmatrix} \gamma_1^* \\ \gamma_m^* \end{pmatrix} = Z \gamma_1^* + Y \gamma_m^* \quad (2.26)$$

y sea V_Z la matriz de proyección engendrada por las columnas de la matriz Z , es decir $V_Z = Z(Z'Z)^{-1}Z'$. Teniendo en cuenta que

$$Y = V_Z Y + (I - V_Z) Y = \frac{1}{S_Y} [Z \hat{a} + e]$$

la combinación lineal de (2.26) se puede poner

$$\begin{aligned} Z^* \gamma_{\min}^* &= Z \gamma_1^* + Y \gamma_m^* = Z \gamma_1^* + \frac{Y_m^*}{S_Y} [Z \hat{a} + e] = \\ &= Z \left[\gamma_1^* + \frac{Y_m^*}{S_Y} \hat{a} \right] + \frac{Y_m^*}{S_Y} e \end{aligned} \quad (2.27)$$

Puesto que e es perpendicular a $C[Z]$, la expresión (2.27) descompone $Z^* \gamma_{\min}^*$ en dos sumandos, cada uno de ellos

en direcciones ortogonales entre si. Por el teorema de Pitágoras

$$\| Z^* \gamma_{\min}^* \|^2 = \| Z \left[\gamma_1^* + \frac{Y_m^*}{S_Y} \hat{\alpha} \right] \|^2 + \frac{Y_m^{*2}}{S_Y^2} \text{SCE} = \lambda_1^*$$

de donde,

$$\| Z \left[\gamma_1^* + \frac{Y_m^*}{S_Y} \hat{\alpha} \right] \|^2 = \lambda_1^* - \frac{Y_m^{*2}}{S_Y^2} \text{SCE} \quad (2.28)$$

Se tiene el siguiente lema auxiliar

LEMA 2.1.

Si los autovalores de $Z'Z$ son todos distintos, entonces

$$\left\| \gamma_1^* + \frac{Y_m^*}{S_Y} \hat{\alpha} \right\| \neq 0$$

DEMOSTRACION:

La demostración se efectua por reducción al absurdo. Si

$$\left\| \gamma_1^* + \frac{Y_m^*}{S_Y} \hat{\alpha} \right\| = 0, \text{ de (2.28) se sigue que } \| Z^* \gamma_{\min}^* \|^2 = \frac{\text{SCE}}{S_Y^2} Y_m^{*2} = \lambda_1^* > 0$$

de modo que, en particular $\gamma_m^* \neq 0$. También

$$\text{si } \left\| \gamma_1^* + \frac{Y_m^*}{S_Y} \hat{\alpha} \right\| = 0 \text{ es } \gamma_1^* + \frac{Y_m^*}{S_Y} \hat{\alpha} = 0 \text{ y de aquí}$$

$$Y_m^* \begin{pmatrix} \hat{\alpha}/S_Y \\ -1 \end{pmatrix} = - \begin{pmatrix} \gamma_1^* \\ Y_m^* \end{pmatrix} = -\gamma_{\min}^*$$

lo que demuestra que $\begin{pmatrix} \hat{\alpha}/S_Y \\ -1 \end{pmatrix}$ es un autovector de Z^*Z^* asociado al autovalor λ_1^* . Por consiguiente,

$$Z^*Z^* \begin{pmatrix} \hat{\alpha}/S_Y \\ -1 \end{pmatrix} = \lambda_1^* \begin{pmatrix} \hat{\alpha}/S_Y \\ -1 \end{pmatrix}$$

o, en forma más desarrollada,

$$\begin{pmatrix} Z'Z & Z'Y \\ Y'Z & 1 \end{pmatrix} \begin{pmatrix} \hat{\alpha}/S_Y \\ -1 \end{pmatrix} = \lambda_1^* \begin{pmatrix} \hat{\alpha}/S_Y \\ -1 \end{pmatrix}$$

es decir

$$\begin{pmatrix} \frac{1}{S_Y} [Z'Z\hat{\alpha} - Z'Y] \\ Y'Z\hat{\alpha} - 1 \end{pmatrix} = \lambda_1^* \begin{pmatrix} \hat{\alpha}/S_Y \\ -1 \end{pmatrix} \quad (2.29)$$

$\hat{\alpha}$ es precisamente solución de las ecuaciones normales $Z'Z\alpha = Z'Y$ de donde (2.29) se convierte en

$$\begin{pmatrix} 0 \\ Y'Z\hat{\alpha} - 1 \end{pmatrix} = \lambda_1^* \begin{pmatrix} \hat{\alpha}/S_Y \\ -1 \end{pmatrix} \quad (2.30)$$

y de aquí $\hat{\alpha} = 0$. Continuando con (2.30), si $\hat{\alpha} = 0$ es $\lambda_1^* = \lambda_{\min}^*[Z^*Z] = 1$. Según el teorema de separación es $\lambda_1 > \lambda_1^* = 1$ lo que conduce a la contradicción $\sum_i \lambda_i > p$.

Una vez demostrado este lema, es fácil establecer el resultado fundamental de este apartado que se enuncia en forma de teorema.

TEOREMA 2.3.

Si los autovalores de $Z'Z$ son todos distintos entre sí, se tiene

$$0 < \lambda_1^* < \lambda_1 \leq \frac{\lambda_1^* - \frac{Y_m^{*2}}{S_Y^2} \text{SCE}}{\left\| Y_1^* + \frac{Y_m^*}{S_Y} \hat{\alpha} \right\|^2} \quad (2.31)$$

DEMOSTRACION

Imponiendo a (2.28) la condición adicional $\left\| Y_1^* + \frac{Y_m^*}{S_Y} \hat{\alpha} \right\| \neq 0$, se sigue que

$$\left\| Z \frac{\left[Y_1^* + \frac{Y_m^*}{S_Y} \hat{\alpha} \right]}{\left\| Y_1^* + \frac{Y_m^*}{S_Y} \hat{\alpha} \right\|} \right\|^2 = \frac{\lambda_1^* - \frac{Y_m^{*2}}{S_Y^2} \text{SCE}}{\left\| Y_1^* + \frac{Y_m^*}{S_Y} \hat{\alpha} \right\|^2} \quad (2.32)$$

Ahora bien, $\lambda_1 = \inf_{c \in \mathbb{R}^1} \|Zc\|^2$ de donde λ_1 es menor o igual que (2.32). Teniendo en cuenta, además, que el teorema de separación establece $0 < \lambda_1^* < \lambda_1$ (c.s.) se llega sin dificultad a (2.31). #

5.2. Comentarios

Las desigualdades (2.31) permiten extraer información del módulo de multicolinealidad de Z a partir de información dependiente únicamente de la matriz Z^* , según la inversa de (2.5). El análisis de (2.31) se facilita si se escribe

$$\frac{SCE}{S_y^2} \gamma_m^{*2} = k_1^* \lambda_1^* \quad (0 < k_1^* < 1)$$

$$\left\| \gamma_1^* + \frac{\gamma_m^*}{S_y} \hat{\alpha} \right\|^2 = k_2^* \lambda_1^* \quad (2.33)$$

de donde (2.31) se puede poner

$$0 < \lambda_1^* < \lambda_1 \leq \frac{1 - k_1^*}{k_2^*} \quad (2.34)$$

En la nueva expresión (2.34), la cota superior depende de las constantes k_1^* y k_2^* definidas en (2.33).

De la condición $\lambda_1^* \leq \lambda_1$ se deduce que si en la matriz Z el módulo es pequeño, en Z^* el módulo es más pequeño todavía. Por otro lado, si el módulo en Z^* es correcto, el módulo en Z también es correcto. Este resultado de agravamiento de la multicolinealidad está en la misma línea que los obtenidos con los VIF_j^* y los números de condición $k[Z]$ y $k[Z^*]$.

Cuando λ_1^* es pequeño, la cuestión queda reducida al análisis de las cantidades k_1^* y k_2^* de (2.33). Una vez conocido que la cota superior de (2.31) es pequeña, como $\lambda_1 = \inf_{c \neq 1} \|Zc\|^2$ es

$$\lambda_1 \leq \left\| Z \frac{\left[\gamma_1^* + \frac{\gamma_m^*}{S_y} \hat{\alpha} \right]}{\left\| \gamma_1^* + \frac{\gamma_m^*}{S_y} \hat{\alpha} \right\|} \right\|^2 \leq \frac{1 - k_1^*}{k_2^*}$$

y así, las coordenadas del vector unitario

$$\frac{\gamma_1^* + \frac{\gamma_m^*}{S_y} \hat{\alpha}}{\left\| \gamma_1^* + \frac{\gamma_m^*}{S_y} \hat{\alpha} \right\|}$$

identifican una dependencia lineal aproximada. En cualquier situación en la que la cota superior de (2.34) sea grande, la desigualdad carece de valor práctico aún con λ_1^* pequeño. Por consiguiente, en este caso, la matriz Z^* no contiene información sobre la multicolinealidad en Z y ésta ha de investigarse directamente.

Finalmente, puede hacerse un comentario sobre la intervención de la respuesta en la relación entre los módulos de Z y Z^* . Intuitivamente, usando ideas de carácter geométrico puede argumentarse que cuanto más cercano a la ortogonalidad esté el vector y con la variedad engendrada por las columnas de Z más depende la estructura de relaciones en la matriz Z^* de lo que

ocurra en Z . En efecto, cuando y tiende a la ortogonalidad con $C[Z]$, la matriz Z^*Z^* tiende a

$$\begin{pmatrix} Z'Z & 0 \\ 0 & 1 \end{pmatrix}$$

cuyos autovalores son los de $Z'Z$ más el valor uno. Por consiguiente, de tener Z^*Z^* un autovalor pequeño éste debe estar entre los autovalores de $Z'Z$. En cierto sentido, esto se refleja en el extremo superior de la cota (2.31) que es función monótona decreciente de SCE que es una medida de la lejanía o proximidad de y a la ortogonalidad de $C[Z]$.

En otro orden de cosas, la estructura de (2.31) indica que el intervalo para λ_1 es ajustado en el caso de multicolinealidades no predictivas. En efecto, en tal situación $\gamma_m^* \approx 0$, $\|\gamma_1^*\| \approx 1$ y, así, el extremo superior de (2.31) está próximo a λ_1^* . En caso de multicolinealidades predictivas, cabe sospechar que la utilización de (2.31) no ha de ser muy eficaz.

Obsérvese que, según (2.5), el cálculo del intervalo

$$\left(\lambda_1^*, \frac{\lambda_1^* - \gamma_m^* \frac{SCE}{S_y^2}}{\|\gamma_1^* + \gamma_m^* \frac{\hat{\alpha}}{S_y}\|^2} \right]$$

depende únicamente del autosistema y la inversa de Z^*Z^* . En efecto, si $(Z^*Z^*)^{-1} = (z^{*jk})$ es

$$\frac{SCE}{S_y^2} = 1 - R^2 = \frac{1}{z^{*mm}}$$

y

$$\frac{\hat{\alpha}}{S_y} = - \left[- \frac{SCE}{S_y^2} \cdot \frac{S_y}{SCE} \hat{\alpha} \right] = - \frac{1}{z^{*mm}} Z^{*(m)} \quad (2.35)$$

donde $Z^{*(m)} = (z^{*1m}, \dots, z^{*pm})$. Por consiguiente, en un problema concreto basta conocer Z^*Z^* para calcular (2.31).

5.3. Ejemplo

Supóngase que se conoce, en una cierta situación, la matriz Z^*Z^* recogida en la tabla 2.1. Es interesante, desde el punto de vista metodológico, ilustrar la aplicación de la desigualdad (2.31)

1	0.934	0.380	0.315
.	1	0.561	0.811
.	.	1	0.513
.	.	.	1

TABLA 2.1. Matriz Z^*Z^*

	$\lambda_1^* = 0.0368$	$\lambda_2^* = 0.2864$	$\lambda_3^* = 0.6777$	$\lambda_4^* = 2.9991$
x_1	0.5941	0.4485	-0.4190	-0.5198
x_2	-0.7788	0.2258	-0.1665	-0.5611
x_3	0.1446	0.1901	0.8901	-0.3882
Y	0.1401	-0.8436	-0.0667	-0.5140

TABLA 2.2a) Autosistema de Z^*Z^*

10.634	-12.006	2.146	1.069
.	16.787	-3.052	-3.514
.	.	1.913	0.031
.	.	.	3.112

TABLA 2.2b) Inversa de Z^*Z

En virtud de (2.35), para este ejemplo se tiene

$$\frac{SCE}{S_y^2} = 1 - R^2 = \frac{1}{2^{*44}} = 0.321$$

$$\frac{\hat{\alpha}}{S_y} = - \frac{1}{2^{*44}} \cdot Z^{*(4)} = \begin{pmatrix} -0.343 \\ 1.128 \\ 0.010 \end{pmatrix}$$

Puesto que $Y_1^* = (0.5941, -0.7788, 0.1446)$, $Y_m^* = 0.1401$, el intervalo de la cota (2.31) informa que

$$\lambda_1^* \in (0.0368, 0.0433] \quad (2.36)$$

Además, el vector $\frac{y_1^* + y_m^* \hat{\alpha} / s_y}{\|y_1^* + y_m^* \hat{\alpha} / s_y\|} = (0.6504, -0.7395, 0.1739)$ identifica una multicolinealidad entre x_1 y x_2 .

El valor verdadero de λ_1 es 0'0433 que coincide con el extremo superior de (2.36). La intención del ejemplo es únicamente ilustrativa. Cuando sólo se pretenda averiguar el valor de λ_1 , es más sencillo calcular el autosistema de $Z'Z$, matriz que coincide con la submatriz principal 3x3 de Z^*Z^* .

6. EJEMPLO

Considérese el siguiente ejemplo tomado del libro de Gunst y Mason (1980) y relativo a un modelo $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$ para los datos de la tabla 2.3. El significado de las variables es

Y = nº de homicidios en cierta ciudad americana por cada 100.00 hab. = HOM.

x_1 = nº de policías por cada 100.000 hab. = FTP.

x_2 = nº de licencias de armas de fuego por cada 100.000 hab. = Lic

x_3 = nº de armas de fuego registradas por cada 100.000 hab. = GR

Las observaciones están tomadas entre los años 1961 y 1973.

Y	x ₁	x ₂	x ₃
8.60	260.35	178.15	215.98
8.90	269.80	156.41	180.48
8.52	272.04	198.02	209.57
8.89	272.96	222.10	231.67
13.07	272.51	301.92	297.65
14.57	261.34	391.22	367.62
21.36	268.89	665.56	616.54
28.03	265.99	1,131.21	1,029.75
31.49	319.87	837.60	786.23
37.39	341.43	794.90	713.77
46.26	356.59	817.74	750.43
47.24	376.69	583.17	1,027.38
52.33	390.19	709.59	666.50

TABLA 2.3

La tabla de análisis de la varianza para un modelo con tres variables y término independiente ajustado con los datos de la tabla 2.3 es

F. Variación	s.c.	g.l.	C.M.	F
Regresión	3,167.41	3	1,055.80	174.80
Error	54.37	9	6.04	
Total (corregido)	3,221.78	12	$R^2 = 0.983$	

n = 13; p = 3; m = 4

TABLA 2.4.

La matriz de correlación ampliada aparece en la tabla 2.5. Destaca inmediatamente la alta correlación entre GR y Lic que podría haberse anticipado del significado de las variables. Por consiguiente, hay una dependencia lineal aproximada entre GR y Lic. Es interesante estudiar los resultados que produce, en cuanto a la detección de multicolinealidad en Z se refiere,

	FTP	Lic	GR	HOM
FTP	1	0.457	0.604	0.931
Lic	.	1	0.904	0.726
GR	.	.	1	0.816
HOM	.	.	.	1

TABLA 2.5. Matriz Z^*Z^*

la utilización de las técnicas de este capítulo.

	FTP	Lic	GR	HOM
FTP	33.338	13.511	3.001	-43.306
Lic	.	10.947	-3.882	-17.363
GR	.	.	8.018	-6.520
HOM	.	.	.	59.256

TABLA 2.6.a) Inversa de Z^*Z^*

	$\lambda_1^* = 0.0106$	$\lambda_2^* = 0.0813$	$\lambda_3^* = 0.6778$	$\lambda_4^* = 3.2303$
FTP	-0.5783	-0.0419	0.6706	-0.4627
Lic	-0.2428	-0.6139	-0.5783	-0.4793
GR	-0.0652	0.7722	-0.3643	-0.5164
HOM	0.7761	-0.1583	0.2882	-0.5381

TABLA 2.6.b) Autosistema de Z^*Z^*

La tabla 2.6.a) recoge la inversa de Z^*Z^* y la tabla 2.6.b) el autosistema. Siguiendo el criterio de (2.11) es

$$\frac{(z^{*14})^2}{z^{*44}} = 31.649 > 27.338 = 33.338 - 6 = z^{*11} - 6$$

$$\frac{(z^{24})^2}{z^{44}} = 5.088 \approx 4.947 = 10.947 - 6 = z^{22} - 6 \quad (2.37)$$

$$\frac{(z^{24})^2}{z^{44}} = 0.717 \leq 2.018 = 8.018 - 6 = z^{33} - 6$$

De (2.37) se deduce que $VIF_{FTP}^* < 6$, $VIF_{Lic}^* \approx 6$ y $VIF_{GR}^* > 6$ lo que demuestra que los datos presentan un ligero grado de colinealidad manifestado en los valores de VIF_{Lic}^* y VIF_{GR}^* .

El cálculo de la desigualdad (2.31) informa de que

$$\lambda_1 \in (0.0106, 0.5671] \quad (2.38)$$

El intervalo (2.38) es poco útil porque sólo con el teorema 2.1 puede afirmarse que $\lambda_1 \in (0.0106; 0.0813)$ lo que indica la existencia de un autovalor pequeño de $Z'Z$. El problema radica en que la multicolinealidad asociada a λ_1^* es predictiva como indica la tabla 2.6.b) y el alto valor de la correlación entre HOM y FTP, 0.931. Como ya se ha apuntado no cabe esperar resultados útiles de (2.31). El vector

$y_1^* + \frac{y_m^*}{S_y} \hat{\alpha}$ que aparece en el cálculo de (2.31) es (0.3971; 0.5560; 0.7292), lo que, de nuevo, apunta hacia las variables GR y Lic. Las cotas que se obtienen como consecuencia del teorema 2.1 son

$$2.8974 < k[Z] < 17.4770$$

$$1 < \eta_2 < 6.3034$$

que no resultan de especial interés.

	VIF_i	$\lambda_1 = 0.0784$	$\lambda_2 = 0.5926$	$\lambda_3 = 2.3290$
FTP	1.690	0.1667	-0.8540	-0.4298
Lic	5.860	0.6471	0.4719	-0.5988
GR	7.301	-0.7440	0.2191	-0.6313

TABLA 2.7.

Finalmente, la tabla 2.7 tiene el interés académico de comprobación de lo obtenido con las técnicas basadas en la matriz

Z^* con lo obtenido con las técnicas clásicas. Efectivamente, los factores de inflación de la varianza de Z detectan a Lic y GR como colineales. Por otro lado, $\lambda_1 = 0.0784$ es pequeño y está asociado a una dependencia lineal aproximada entre GR y Lic que son las variables correspondientes a las coordenadas grandes de Y_1 . Es de destacar que los autovalores de $Z'Z$ y $Z^{*'}Z^*$ satisfacen el teorema 2.1.

CAPITULO III

1. SUMARIO

Este capítulo contiene el desarrollo de un nuevo método de detección simultánea de multicolinealidad y casos influyentes.

La sección 2 se dedica a la exposición de las ideas básicas. En breve, se trata de considerar, para distintos diagnósticos, la matriz resultante de añadir a la matriz de diseño, ampliada o no, la columna formada por el i -ésimo vector canónico de \mathbb{R}^n . Se ilustra cómo la condición de la matriz perturbada puede informar de la condición de la matriz original y del potencial del caso i -ésimo.

Por razones de sencillez en la exposición, es conveniente empezar, en la sección 3, por el método reducido de detección únicamente de multicolinealidad y filas extremas. Dicho método, se deduce como consecuencia de ciertos resultados teóricos relativos al comportamiento de los diagnósticos de multicolinealidad de la matriz perturbada. La sección termina con la aplicación del método a un ejemplo práctico.

La sección 4 incluye el método completo que, en pocas palabras, se puede considerar una generalización de los resultados de la sección 3. El esquema es esencialmente el mismo: tras los desarrollos teóricos, el método se obtiene como consecuencia directa. Como se demuestra, la aplicación del método necesita de ciertos resultados del Capítulo II. El mismo ejemplo utilizado en la sección 3, es el punto final de este capítulo.

2. MOTIVACION

Hocking y Dunn (1982) señalan que, tradicionalmente, los problemas de multicolinealidad y casos extremos de la matriz de diseño se tratan por separado. Hablando en términos informales, la multicolinealidad se referiría a problemas con las columnas y los casos extremos a problemas con las filas de la matriz X . En un intento de analizar ambos problemas conjuntamente, Hocking y Dunn proponen los gráficos de componentes principales, diagnóstico de importante aplicación práctica que aparecerá posteriormente en este trabajo. En concreto, el propósito de este capítulo es desarrollar una estructura que también permite el aná-

lisis simultáneo de ambos problemas. El método se puede extender fácilmente a la detección simultánea de multicolinealidad y casos influyentes en el modelo (1.1).

La idea es sencilla. Considérese la matriz de $n \times m$

$$Z_i = [Z \mid d_i] \quad (3.1)$$

donde $d_i = (0, \dots, 1, \dots, 0)'$ es el i -ésimo vector canónico de \mathbb{R}^n . Es intuitivo que al ser Z_i una ligera perturbación de Z , las condiciones de Z_i y Z han de estar estrechamente relacionadas. Por otro lado, sea $V_Z = Z(Z'Z)^{-1}Z'$ la matriz de proyección sobre $C[Z]$. La distancia de d_i a la variedad $C[Z]$ es la distancia entre los vectores d_i y $V_Z d_i$. Esta distancia al cuadrado es

$$\|d_i - V_Z d_i\|^2 = d_i' [I - V_Z] d_i = 1 - \sigma_{ii}[Z] \quad (3.2)$$

dada la idempotencia de la matriz V_Z . En (3.2) $\sigma_{ii}[Z] = z_i'(Z'Z)^{-1}z_i$, el i -ésimo elemento diagonal de la matriz V_Z , cumple la relación $\sigma_{ii} = \frac{1}{n} + \sigma_{ii}[Z]$ de modo que cuanto más potencial tiene el caso i -ésimo, menor es la distancia de d_i a $C[Z]$. Si se escribe $V_Z d_i = Z c_i$, donde c_i es $p \times 1$, (3.2) se puede poner

$$\|d_i - Z c_i\|^2 = \|d_i - V_Z d_i\|^2 = 1 - \sigma_{ii}[Z] \quad (3.3)$$

y por consiguiente el que x_i' sea extrema provoca la aparición de una dependencia lineal aproximada entre d_i y Z y, con ello, la mala condición de la matriz Z_i . En suma cabe pensar que el análisis de la condición de las matrices Z_i ($i=1, \dots, n$) ha de contener toda la información necesaria para el conocimiento de la condición de Z y la distancia de las filas de la matriz de diseño al centro de los datos. El apartado 3 de este capítulo es la respuesta afirmativa a esta conjetura.

Una generalización inmediata de (3.1) es la matriz

$$Z_i^* = [Z^* \mid d_i] \quad (3.4)$$

que contiene información de la multicolinealidad en Z^* y, según el capítulo 2, de la multicolinealidad en Z . El análogo de (3.3) es

$$\|d_i - Z^* c_i^*\|^2 = 1 - \sigma_{ii}^*[Z] \quad (3.5)$$

donde $\sigma_{ii}^* = \frac{1}{n} + \sigma_{ii}^*[z]$, $\sigma_{ii}^*[z] = z_i^{*'}(Z^{*'}Z^*)^{-1}z_i^*$ y $V_{2^*}d_i = Z^{*'}c_i^*$. Tomando σ_{ii}^* como medida de la influencia de un caso, la fórmula (3.5) sugiere que un caso influyente provoca problemas de degeneración en la matriz (3.4) asociada. La condición de las matrices Z_i^* ($i=1, \dots, n$) es, pues una herramienta de análisis conjunto de multicolinealidad y casos influyentes.

De modo general, el capítulo está dividido en dos secciones diferenciadas, dedicadas, respectivamente, a las matrices Z_i y Z_i^* . A su vez, cada sección se descompone en dos partes: una dedicada a un estudio de carácter teórico y otra destinada al desarrollo de las aplicaciones de los resultados obtenidos.

3. ANALISIS DE MULTICOLINEALIDAD EN Z_i

Razones de tipo metodológico aconsejan empezar los desarrollos por el análisis de las matrices Z_i ($i=1, 2, \dots, n$).

3.1. Factores de inflación de la varianza

Sea $i \in \{1, 2, \dots, n\}$ fijo. La matriz de productos cruzados asociada a la matriz correspondiente de (3.1) es

$$Z_i'Z_i = \begin{pmatrix} Z'Z & z_i' \\ z_i & 1 \end{pmatrix}$$

siendo z_i' la i -ésima fila de la matriz Z . Usando el resultado de inversión adecuado, la inversa de (3.6) es

$$(Z_i'Z_i)^{-1} = \begin{pmatrix} [Z'Z - z_i z_i']^{-1} & -\frac{1}{1 - \sigma_{ii}[z]} a_i' \\ -\frac{1}{1 - \sigma_{ii}[z]} a_i & \frac{1}{1 - \sigma_{ii}[z]} \end{pmatrix} \quad (3.7)$$

donde $a_i = (Z'Z)^{-1} z_i$. Además,

$$[Z'Z - z_i z_i']^{-1} = (Z'Z)^{-1} + \frac{a_i a_i'}{1 - \sigma_{ii}[z]} \quad (3.8)$$

Si se denotan por $VIF_{j,i}$ ($j=1, \dots, p$) los p primeros elementos diagonales de la matriz $(Z_i'Z_i)^{-1}$ y por VIF^i el último, se tiene, según (3.7) y (3.8),

$$VIF_{j,i} = VIF_j + \frac{a_{j,i}^2}{1 - \sigma_{ii}[z]} \quad (j=1,2,\dots,p) \quad (3.9)$$

y

$$VIF^i = \frac{1}{1 - \sigma_{ii}[z]} \quad (3.10)$$

Un primer resultado es que la inversa de $Z_i' Z_i$ permite calcular $\sigma_{ii}[z]$ y, de aquí, σ_{ii} . Conocido el vector a_i , de (3.9) se pueden calcular los VIF_j ($j=1,2,\dots,p$).

Sea $c_j = (0, \dots, 1, \dots, 0)$ el j -ésimo vector canónico de \mathbb{R}^p . En virtud de la desigualdad de Cauchy-Schwartz es

$$\begin{aligned} a_{j,i}^2 &= [c_j'(Z'Z)^{-1} z_i]^2 \leq [c_j'(Z'Z)^{-1} c_j] [z_i'(Z'Z)^{-1} z_i] = \\ &= VIF_j \cdot \sigma_{ii}[z] \end{aligned} \quad (3.11)$$

Por consiguiente si se escribe

$$VIF_{j,i} = VIF_j + \frac{a_{j,i}^2}{1 - \sigma_{ii}[z]} = VIF_j \left[1 + \frac{a_{j,i}^2 / VIF_j}{1 - \sigma_{ii}[z]} \right]$$

se deduce que el segundo sumando en el corchete está acotado por $\sigma_{ii}[z] [1 - \sigma_{ii}[z]]^{-1}$ y, así,

$$VIF_{j,i} \leq VIF_j \frac{1}{1 - \sigma_{ii}[z]} = VIF_j \cdot VIF^i, \quad j=1,\dots,p \quad (3.12)$$

En resumen, de (3.9) y (3.12) se sigue que

$$VIF_j \leq VIF_{j,i} \leq \frac{VIF_j}{1 - \sigma_{ii}[z]}, \quad j=1,\dots,p \quad (3.13)$$

Según (3.13), es inmediato que, $VIF_{j,i} \rightarrow VIF_j$ cuando $\sigma_{ii}[z] \rightarrow 0$ ($j=1,\dots,p$). Esta condición puede obtenerse como corolario del siguiente teorema.

TEOREMA 3.1.

Si $\sigma_{ii}[z] \rightarrow 0$, la norma euclídea de la diferencia entre $(Z_i' Z_i)^{-1}$ y la matriz $m \times m$

$$\begin{pmatrix} (Z'Z)^{-1} & 0 \\ 0 & 1 \end{pmatrix} \quad (3.14)$$

tiende a cero.

DEMOSTRACION:

Si $\|\cdot\|_E$ denota la norma euclídea de una matriz es fácil ver, según (3.7), (3.8) y (3.11), que

$$\left\| \begin{pmatrix} (Z_i' Z_i)^{-1} & 0 \\ 0 & 1 \end{pmatrix} \right\|_E \leq \frac{\sigma_{ii}[Z]}{1 - \sigma_{ii}[Z]} + \sigma_{ii}[Z] \left[\sum_{j=1}^p V_i F_j + \sum_{j \neq k} V_i F_j V_i F_k \right] \dots$$

cuando $\sigma_{ii}[Z] \rightarrow 0$ #

Un comentario evidente a este resultado es que, en relación con (3.14), la perturbación causada en la inversa de la matriz de productos cruzados al añadir la columna d_i a la matriz de diseño Z , es tanto menor cuanto más pequeño sea $\sigma_{ii}[Z]$. Este hecho será posteriormente utilizado en las aplicaciones prácticas de esta teoría.

3.2. Autovalores e índices de condición

Sean $0 \leq \lambda_{1,i} \leq \lambda_{2,i} < \dots < \lambda_{p,i} < \lambda_{m,i}$ los autovalores de la matriz $Z_i' Z_i$. Es fácil probar que la matriz Z_i es de rango completo. En efecto, la matriz Z es de rango p y si $d_i \in C[Z]$, entonces $d_i = V_2 d_i$ y según (3.3) $\sigma_{ii}[Z] = 1$. Pero $\sigma_{ii} \leq 1$ de modo que $\sigma_{ii}[Z] \leq 1 - \frac{1}{n}$ y, por consiguiente, el vector d_i no puede depender linealmente de las columnas de Z con lo que la matriz Z_i es de rango $p+1=m$. Así pues, la matriz $Z_i' Z_i$ es definida positiva y $\lambda_{1,i} > 0$.

Se tiene un teorema de separación similar al teorema 2.1:

TEOREMA 3.2.

Sea C la matriz ortogonal $p \times p$ de la descomposición espectral de la matriz $Z' Z$. Si los autovalores de $Z' Z$ son todos distintos y $k_i = C' Z_i$ es un vector $p \times 1$ cuyos componentes son todos distintos de cero, se tiene

$$0 < \lambda_{1,i} < \lambda_1 < \lambda_{2,i} < \dots < \lambda_p < \lambda_{m,i}$$

DEMOSTRACION:

La técnica de la demostración es similar a la del teorema 2.1. Los polinomios característicos de $Z_i' Z_i$ y de $\tilde{C}' Z_i' Z_i \tilde{C}$ son idénticos, si \tilde{C} está dada por

$$\tilde{C} = \begin{pmatrix} C & 0 \\ 0 & 1 \end{pmatrix}$$

Pero

$$\tilde{C}' Z_i' Z_i \tilde{C} = \begin{pmatrix} 0 & k_i \\ k_i' & 1 \end{pmatrix}$$

y análogamente al teorema 2.1, las raíces del polinomio característico de $Z_i' Z_i$ coinciden con las de la ecuación

$$1 = \lambda_{.i} + \sum_{j=1}^p \frac{k_{ij}^2}{\lambda_j - \lambda_{.i}}$$

donde $k_i' = (k_{i1}, \dots, k_{ip})$ y $\lambda_{.i}$ es un autovalor genérico de $Z_i' Z_i$. Una figura similar a la fig. 2.1 conduce a la conclusión del enunciado. #

Una consecuencia inmediata de este teorema es que, medida por el menor de los autovalores, la multicolinealidad de Z_i es más grave que la de Z . Por otro lado, pueden establecerse desigualdades entre los índices de condición de las matrices Z y Z_i . Sean

$$\eta_k = \left[\frac{\lambda_{\max}}{\lambda_k} \right]^{1/2} \quad , \quad \eta_{k,i} = \left[\frac{\lambda_{\max,i}}{\lambda_{k,i}} \right]^{1/2}$$

($k = 1, \dots, p$) ($k = 1, \dots, p, m$)

y

$$\delta_{k,i} = \left[\frac{\lambda_{p,i}}{\lambda_{k+1,i}} \right]^{1/2} \quad (k = 1, \dots, p-1)$$

Usando el teorema de separación 3.1 es inmediato que

$$\delta_{k,i} < \eta_k < \eta_{k,i} \quad , \quad k = 1, \dots, p-1 \quad (3.15)$$

Tomando $k=1$ en (3.15) se deduce para el número de condición

$$\delta_{1,i} < k[Z] < k[Z_i] \quad (3.16)$$

Como ya ocurría en el capítulo II, los resultados obtenidos con el autosistema no son tan precisos como los obtenidos con los factores de inflación de la varianza en el apartado 3.1. No obstante, puede establecerse el siguiente teorema, complementario del teorema 3.1.

TEOREMA 3.3.

Cuando $\sigma_i[z] \rightarrow 0$, $[\lambda_{1,i}, \dots, \lambda_{m,i}] \rightarrow [\lambda_1, \dots, \lambda_p, 1]$ donde $[x, \dots, y]$ indica una ordenación de coordenadas en sentido creciente.

DEMOSTRACION:

En virtud del teorema 3.2, la inversa de $Z_i' Z_i$ tiende en norma euclídea, a la matriz (3.14). Los autovalores de esta matriz son $[\lambda_{1,i}^{-1}, \dots, \lambda_{p,i}^{-1}, 1]$ y, por consiguiente, teniendo en cuenta la continuidad de los autovalores considerados como funciones de la matriz correspondiente (Kato (1982), pág. 136), se tiene que

$$[\lambda_{1,i}^{-1}, \dots, \lambda_{m,i}^{-1}] \rightarrow [\lambda_1^{-1}, \dots, \lambda_p^{-1}, 1]$$

de donde $[\lambda_{1,i}, \dots, \lambda_{m,i}] \rightarrow [\lambda_1, \dots, \lambda_p, 1]$ por tratarse de autovalores de matrices definidas positivas.

De este resultado, se deduce que $\lambda_{1,i} \rightarrow \lambda_1$, $\lambda_{m,i} \rightarrow \lambda_p$ y, por consiguiente, $k^2[Z_i] \rightarrow k^2[Z]$ cuando $\sigma_i[z] \rightarrow 0$. La importancia práctica de esta conclusión será estudiada posteriormente.

Hay que observar que el teorema 3.3 tiene un inmediato análogo para autovectores (ver Kato (1982), pág. 136).

3.3.1. Módulo de Gunst

En este apartado es necesario comenzar probando el siguiente

LEMA 3.1.

Sean V_2 la matriz de proyección sobre la variedad $C[Z]$, d_i el i -ésimo vector canónico de \mathbb{R}^n y z_i' la i -ésima fila de la matriz Z ($i=1, \dots, n$). Si $V_2 d_i = Z c_i$, donde c_i es $p \times 1$, y $\gamma_{\min,i} = (\gamma_{1,i}, \gamma_{m,i})'$ autovector normalizado asociado a $\lambda_{1,i}$, entonces $z_i \neq 0$ implica que $\gamma_{1,i} + \gamma_{m,i} c_i \neq 0$ ($i=1, \dots, n$).

DEMOSTRACION:

Se trata de llegar a una contradicción cuando $\gamma_{1,i} + \gamma_{m,i} c_i = 0$. Si se pone

$$Z_i \gamma_{\min,i} = \begin{bmatrix} Z & d_i \end{bmatrix} \begin{pmatrix} \gamma_{1,i} \\ \gamma_{m,i} \end{pmatrix} = Z \gamma_{1,i} + \gamma_{m,i} d_i$$

y se tiene en cuenta que

$$d_i = V_2 d_i + (I - V_2) d_i = Z c_i + (I - V_2) d_i$$

entonces

$$Z_i \gamma_{\min,i} = Z [\gamma_{1,i} + \gamma_{m,i} c_i] + \gamma_{m,i} [I - V_2] d_i$$

expresión que depende de los vectores ortogonales

$Z [\gamma_{1,i} + \gamma_{m,i} c_i]$ y $[I - V_2] d_i$. Usando el teorema de Pitágoras

$$\lambda_{1,i} = \| Z_i \gamma_{\min,i} \|^2 = \| Z [\gamma_{1,i} + \gamma_{m,i} c_i] \|^2 + \gamma_{m,i}^2 [1 - \sigma_{ii}[z]] \quad (3.17)$$

La identidad (3.17) permite deducir, en primer lugar, $\lambda_{1,i} = \gamma_{m,i}^2 [1 - \sigma_{ii}[z]] > 0$ de donde $\gamma_{m,i} \neq 0$ y, de aquí, en segundo lugar

$$\gamma_{m,i} \begin{pmatrix} c_i \\ -1 \end{pmatrix} = \begin{pmatrix} \gamma_{m,i} c_i \\ -\gamma_{m,i} \end{pmatrix} = - \begin{pmatrix} \gamma_{1,i} \\ \gamma_{m,i} \end{pmatrix}$$

lo que demuestra que $(c_i', -1)'$ es un autovector de $Z_i' Z_i$ asociado al autovalor $\lambda_{1,i}$. En forma desarrollada esto último puede escribirse

$$\begin{aligned} Z_i' Z_i \begin{pmatrix} c_i \\ -1 \end{pmatrix} &= \begin{pmatrix} Z' Z & z_i \\ z_i' & -1 \end{pmatrix} \begin{pmatrix} c_i \\ -1 \end{pmatrix} = \\ &= \begin{pmatrix} Z' Z c_i - z_i \\ z_i' c_i - 1 \end{pmatrix} = \lambda_{1,i} \begin{pmatrix} c_i \\ -1 \end{pmatrix} \end{aligned} \quad (3.18)$$

Como $V_2 d_i = Z c_i$, se tiene

$$Z' Z c_i = Z' V_2 d_i = Z' d_i = z_i$$

y

$$z_i' c_i = d_i' Z c_i = d_i' V_2 d_i = \sigma_{ii}[z]$$

de modo que sustituyendo en (3.18)

$$\begin{pmatrix} 0 \\ -[1 - \sigma_{ii}[z]] \end{pmatrix} = \lambda_{1,i} \begin{pmatrix} c_i \\ -1 \end{pmatrix}$$

lo que implica $c_i = 0$ y $\lambda_{1,i} = 1 - \sigma_{ii}[z]$.

Pero si $c_i = 0$ es que $V_2 d_i = Z c_i = 0$ lo que, en virtud

de la idempotencia de la matriz V_2 , es equivalente a que $\sigma_i[z] = \|V_2 d_i\|^2 = 0$ y, a su vez, esto implica que $z_i = 0$ por ser $(Z'Z)^{-1}$ una matriz definida positiva. La condición $z_i = 0$ es equivalente a $x_{1i} = \bar{x}$, situación de rara aparición práctica. #

Una vez demostrado este lema, el principal resultado de este apartado es el siguiente

TEOREMA 3.4.

Si $z_i \neq 0$, se tiene

$$0 < \lambda_{1,i} < \lambda_1 \leq \frac{\lambda_{1,i} - \gamma_{m,i}^2 [1 - \sigma_i[z]]}{\| \gamma_{1,i} + \gamma_{m,i} c_i \|^2} \quad (3.19)$$

DEMOSTRACION:

Puesto que $\lambda_1 = \min_{c'c=1} \|Zc\|^2$, es, de (3.17),

$$\lambda_1 \leq \left\| Z \frac{[\gamma_{1,i} + \gamma_{m,i} c_i]}{\| \gamma_{1,i} + \gamma_{m,i} c_i \|} \right\|^2 = \frac{\lambda_{1,i} - \gamma_{m,i}^2 [1 - \sigma_i[z]]}{\| \gamma_{1,i} + \gamma_{m,i} c_i \|^2}$$

y usando el teorema de separación 3.2 del apartado anterior se llega a la desigualdad

$$0 < \lambda_{1,i} < \lambda_1 \leq \frac{\lambda_{1,i} - \gamma_{m,i}^2 [1 - \sigma_i[z]]}{\| \gamma_{1,i} + \gamma_{m,i} c_i \|^2} \quad \#$$

La interpretación de (3.19) se facilita, como ocurría con (2.31), si se pone

$$\begin{aligned} \gamma_{m,i}^2 [1 - \sigma_i[z]] &= k_{1,i} \lambda_{1,i} \\ \| \gamma_{1,i} + \gamma_{m,i} c_i \|^2 &= k_{2,i} \lambda_{1,i} \end{aligned} \quad (3.20)$$

con lo cual (3.19) se puede escribir

$$0 < \lambda_{1,i} < \lambda_1 \leq \frac{1 - k_{1,i}}{k_{2,i}} \quad (3.21)$$

Las cantidades $k_{1,i}$ y $k_{2,i}$ de (3.20) dependen únicamente del auto-sistema de la matriz $Z_i' Z_i$ y, en conjunto, componen una cota superior para el módulo de Gunst de Z . Si esta cota superior es pequeña, el módulo $\lambda_1^{1/2}$ es pequeño, y por (3.17), las coorde-

nadas del vector

$$\frac{\gamma_{1,i} + \gamma_{m,i} c_i}{\|\gamma_{1,i} + \gamma_{m,i} c_i\|}$$

identifican una dependencia lineal aproximada en Z . Según (3.21), la investigación de la multicolinealidad es innecesaria cuando $\lambda_{1,i}^{1/2}$, el módulo de Z_i , tenga un valor correcto. La desigualdad (3.21) no es de utilidad cuando $\lambda_{1,i}$ es pequeño y la cota superior $(1-k_{1,i})/k_{2,i}$ es alta.

Usando el análogo, del teorema 3.3 para autovectores, cuando $\sigma_{ii}[z] \rightarrow 0$ $\gamma_{\min,i} \rightarrow (\gamma_1, 0)$ en virtud de la estructura de (3.14). Por consiguiente, el extremo superior de (3.19) tiende a λ_1 cuando el potencial decrece hacia cero.

3.4. Desigualdades de Berk

El teorema del apartado 4.3 es adaptable, con ligeras variantes, a la matriz $Z_i' Z_i$. Dada la estructura de

$$Z_i' Z_i = \begin{pmatrix} Z_i' Z_i & z_i \\ z_i' & 1 \end{pmatrix}$$

la técnica de la demostración utilizada en 4.3 puede utilizarse para obtener

TEOREMA 3.5.

Se cumple

$$\max \{VIF^i, VIF_{1,i}, \dots, VIF_{p,i}\} \leq k^2[Z_i] \leq m \left[VIF^i + \sum_{j=1}^p VIF_{j,i} \right]$$

DEMOSTRACION:

La idea de la demostración es que al ser $\sum_{j=1}^m \lambda_{j,i} = m$ entonces $1 \leq \lambda_{m,i} \leq m$. Por consiguiente,

$$\max \{VIF^i, VIF_{1,i}, \dots, VIF_{m,i}\} \leq \max_{c \geq 1} c'(Z_i' Z_i)^{-1} c = \frac{1}{\lambda_{1,i}}$$

y sin más que usar el hecho de ser $\lambda_{m,i} \geq 1$, se llega a la acotación inferior. Por otro lado, se tiene

$$\frac{1}{\lambda_{1,i}} \leq \text{tr}[(Z_i' Z_i)^{-1}] = VIF^i + \sum_{j=1}^p VIF_{j,i} \quad (3.22)$$

de donde, multiplicando (3.22) por $\lambda_{m,i}$ ($\leq m$) se llega a la acotación superior. #

Por su interés práctico, se incluye la demostración de un

lema útil en aplicaciones.

LEMA 3.2.

$$k^2[Z] < k^2[Z_i] \leq \frac{m}{1 - \sigma_{ii}[Z]} [1 + \rho k^2[Z]] \quad (3.23)$$

DEMOSTRACION:

Basta demostrar únicamente la cota superior. En virtud de la desigualdad de Cauchy-Schwartz es $VIF_{j,i} \leq VIF_j \cdot VIF_i^L$ según (3.11). Puesto que $VIF_i^L = [1 - \sigma_{ii}[Z]]^{-1}$, se tiene

$$VIF_i^L + \sum_{j=1}^p VIF_{j,i} \leq VIF_i^L \left[1 + \sum_{j=1}^p VIF_j \right] = \frac{1}{1 - \sigma_{ii}[Z]} \left[1 + \sum_{j=1}^p VIF_j \right] \quad (3.24)$$

Aplicando el teorema de Berk a la matriz de correlación $C = -Z'Z$ es $\sum_{j=1}^p VIF_j \leq p \max_j VIF_j \leq p k^2[Z]$ de donde siguiendo en (3.24):

$$VIF_i^L + \sum_{j=1}^p VIF_{j,i} \leq \frac{1}{1 - \sigma_{ii}[Z]} \left[1 + \sum_{j=1}^p VIF_j \right] \leq \frac{1}{1 - \sigma_{ii}[Z]} [1 + \rho k^2[Z]] \quad (3.25)$$

Usando el teorema anterior y (3.25) es, finalmente,

$$k^2[Z_i] \leq \frac{m}{1 - \sigma_{ii}[Z]} [1 + \rho k^2[Z]] \quad \#$$

3.5. Aplicaciones: detección simultánea de colinealidad y potencial

Los resultados de los apartados anteriores, permiten establecer un método de análisis combinado de colinealidad y "leverage". El método podrá esquematizarse en las dos etapas siguientes:

i) Asociado a cada caso i , se calcula la magnitud VIF_i^L de (3.10). Si VIF_i^L es alto, la fila i -ésima de la matriz de diseño es extrema. Teniendo en cuenta que un punto de corte para $\sigma_{ii}[Z]$ es $\sigma_{ii}^{(0)} > 2p/n$, un punto de corte para VIF_i^L puede ser

$$VIF_i^L > \frac{n}{n - (2p - 1)}$$

Cuando la dimensión del problema es baja (m pequeño) el cálculo de las n cantidades VIF_i^L ($i=1, \dots, n$) podría ejecutar-

se empleando por ejemplo subrutinas del paquete de programas IMSL.

ii) Sea i_0 tal que $\min VIF^i = VIF^{i_0}$. En virtud de los teoremas 3.1 y 3.3, si $\sigma_{i_0 i_0}^2[Z] \approx 0$ o, equivalentemente, $VIF^{i_0} \approx 1$ los diagnósticos de multicolinealidad en $Z_{i_0} Z_{i_0}$ son buenas aproximaciones de los diagnósticos de multicolinealidad en $Z'Z$. Por consiguiente, la información deducida de las cantidades VIF_{j,i_0}^i ($j=1, \dots, p$), λ_{i,i_0} y $k^2[Z_{i_0}]$ es suficiente para el análisis de multicolinealidad en Z . Cuando $\sigma_{i_0 i_0}^2[Z]$ no sea suficientemente pequeño, las expresiones (3.9) y las desigualdades (3.13), (3.19) y (3.2) pueden ser de utilidad, aplicadas en $i=i_0$.

3.6. Ejemplo

La teoría de este apartado puede ilustrarse con un ejemplo muy conocido que ha sido analizado por Cook-Weisberg (1982), Weisberg (1985) y Cook (1986), entre otros. Se trata de un modelo $Y = \beta_0 + \beta_1 x_1 + \dots + \beta_3 x_3 + \varepsilon$ sobre $n = 19$ ratas y con variables

Y = Porcentaje retenido de una dosis de cierta droga en el hígado.

x_1 = peso de la rata (g.)

x_2 = peso del hígado (g.)

x_3 = dosis relativa de droga administrada.

Los datos aparecen recogidos en la tabla 3.1.

Y	x ₁	x ₂	x ₃	
0.42	176	6.5	0.88	
0.25	176	9.5	0.88	
0.56	190	9.0	1.00	
0.23	176	8.9	0.88	
0.23	200	7.2	1.00	
0.32	167	8.9	0.83	
0.37	188	8.0	0.94	
0.41	195	10.0	0.98	
0.33	176	8.0	0.88	
0.38	165	7.9	0.84	
0.27	158	6.9	0.80	
0.36	148	7.3	0.74	
0.21	149	5.2	0.75	
0.28	163	8.4	0.81	
0.34	170	7.2	0.85	n = 19
0.28	186	6.8	0.94	p = 3
0.30	146	7.3	0.73	m = 4
0.37	181	9.0	0.90	
0.46	149	6.4	0.75	

TABLA 3.1.

Para este modelo existe una multicolinealidad estructural porque una de las reglas bajo las que se realiza el experimento es

$$x_3 \approx k \frac{40 \text{ mg}}{1000} x_1$$

es decir, la dosis administrada es 40 mg. de droga por Kg. de peso corporal. Esto se confirma plenamente por la matriz de correlaciones muestrales

$$Z'Z = \begin{pmatrix} 1 & - & - \\ 0.5 & 1 & \\ 0.99 & 0.49 & 1 \end{pmatrix}$$

Es de destacar que, en los estudios anteriores, esta cuestión no ha sido analizada. Además, en este ejemplo hay un caso, el núme-

ro 3, que con $\psi_{ii} = 0.85$, es influyente y distorsiona los resultados del análisis mínimo-cuadrático. Ambos problemas, multicolinealidad y potencial, pueden estudiarse con los métodos de este apartado.

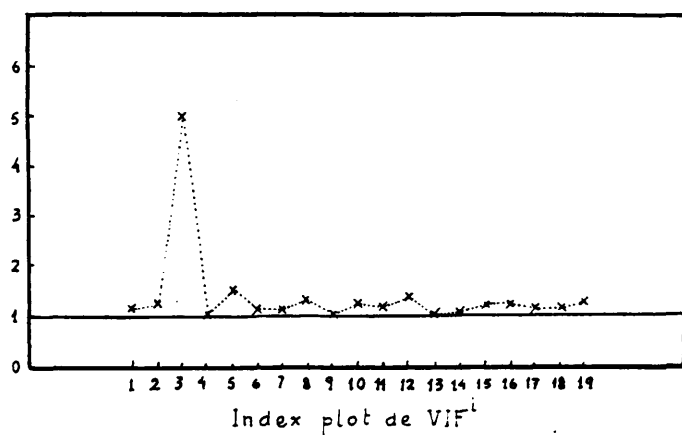


Fig. 3.1.

La figura 3.1 es un "index plot" de los VIF^i . Se detecta inmediatamente como destacado el caso número 3 con $VIF^3 = 4.959$. Una información de interés en lo que sigue es que $\min_i VIF^i = VIF^1 = 1.015$. ($\psi_{qq}[2] = 0.014$).

La tabla 3.2 a) recoge los factores de inflación de la varianza para $Z_q' Z_q$ así como el número de condición al cuadrado de esta matriz. Por su parte, la tabla 3.2.b) hace lo propio con los pasos intermedios necesarios para el cálculo de la desigualdad (3.19).

i_0	$VIF_{1,q}$	$VIF_{2,q}$	$VIF_{3,q}$	$K^2[Z_q]$
9	51.3974	1.3356	50.664	240.611

Tabla 3.2.a)

i_0	$\lambda_{1,q}$	$R_{1,q}^2 [1-u_{1,q}(a)]$	$\ r_{1,q} + r_{2,q} c_q\ ^2$
9	0.0098	0.00113	0.9772

Tabla 3.2.b)

Como $\sigma_{qq}[Z] \approx 0$, el teorema 3.1 garantiza que $VIF_{j,q} \approx VIF_j$ ($j=1,2,3$). La tabla 3.2.a) identifica inmediatamente como colineales las variables x_1 y x_3 . Del mismo modo, el teorema 3.3 implica que $\lambda_{1,q} \approx \lambda_1$. No obstante, el método de la desigualdad (3.19) permite deducir que

$$\lambda_1 \in (0.0098, 0.0099]$$

lo que implica que λ_1 es pequeño. El vector de multicolinealidad $\frac{Y_{1,q} + Y_{n,q} c_q}{\|Y_{1,q} + Y_{n,q} c_q\|}$ es (0.7095, -0.0095, -0.7047) lo que indica que hay una dependencia lineal aproximada entre x_1 y x_3 . Es de destacar la gran precisión de los resultados obtenidos mediante las aproximaciones de esta sección.

Finalmente, la tabla 3.3 tiene el interés de comprobar lo obtenido con las técnicas clásicas de investigación de multicolinealidad.

	VIF_j	$\lambda_1 = 0.0099$	$k^2[Z] = 236.88$
x_1	51.00	-0.7095	
x_2	1.3356	-0.0095	
x_3	50.3356	-0.7047	

TABLA 3.3.

3.7. Desarrollos sin tipificar los datos

Los desarrollos de esta sección se pueden efectuar sin necesidad de tipificar las variables regresoras utilizando la matriz

$$X_i = [X \mid d_i] \quad (3.26)$$

de orden $n \times (m+1)$. En general, los resultados que se obtienen son similares a los de la matriz Z_i desde un punto de vista formal aunque, no obstante, el significado de los mismos pueda ser bastante discutible. La ausencia de tipificación implica la inexistencia de un marco común de escalas en las columnas de la matriz X_i y, por consiguiente, no es muy correcto hablar de mala condición al contrario que en (3.1) desde las columnas son todas de longitud unitaria.

Sin embargo, la matriz $(X_i' X_i)^{-1}$ contiene información interesante. Usando el resultado de inversión usual se sigue que

$$(X_i' X_i)^{-1} = \left(\begin{array}{c|c} \frac{(X'X)^{-1} (X'X)^{-1} x_i x_i' (X'X)^{-1}}{1 - \sigma_{ii}} & - \frac{(X'X)^{-1} x_i}{1 - \sigma_{ii}} \\ \hline - \frac{x_i' (X'X)^{-1}}{1 - \sigma_{ii}} & \frac{1}{1 - \sigma_{ii}} \end{array} \right) = \quad (3.27)$$

$$= \left(\begin{array}{c|c} (X_{(i)}' X_{(i)})^{-1} & - \frac{(X'X)^{-1} x_i}{1 - \sigma_{ii}} \\ \hline - \frac{x_i' (X'X)^{-1}}{1 - \sigma_{ii}} & \frac{1}{1 - \sigma_{ii}} \end{array} \right)$$

Por consiguiente, la inversa de $X_i' X_i$ permite deducir:

i) En notación obvia, $VIF_i = \frac{1}{1 - \sigma_{ii}}$ y de aquí el valor de la distancia σ_{ii} .

ii) Además

$$VIF_{j,i} = VIF_{j(i)} = VIF_j + \frac{\alpha_j^2}{1 - \sigma_{ii}} \quad (3.28)$$

$$j = 1, 2, \dots, p, p+1 \quad , \quad i = 1, \dots, n$$

donde $\alpha = (\alpha_1, \dots, \alpha_p, \alpha_m)' = (X'X)^{-1} x_i$. Según (3.28) los primeros elementos diagonales de la matriz $(X_i' X_i)^{-1}$ son los m factores de inflación de la varianza calculados eliminando el caso i -ésimo del análisis.

iii) La parte no aleatoria del cambio

$$\hat{\beta} - \hat{\beta}_{(i)} = \frac{(X'X)^{-1}X_i}{1 - \sigma_{ii}} e_i \quad (i = 1, 2, \dots, n)$$

en los m primeros elementos de la última columna de $(X_i'X_i)^{-1}$.

Finalmente, la magnitud σ_{ii} podrá interpretarse como una distancia de d_i a la variedad engendrada por las columnas de la matriz X . Una diferencia importante con (3.3) es que, en la representación, con notación evidente

$$\|d_i - Xc_i\|^2 = 1 - \sigma_{ii}$$

puede darse, al menos en teoría, la situación $\sigma_{ii} = 1$. Por tanto, la matriz X_i puede no ser de rango completo al contrario que Z_i que siempre lo es. Si $\sigma_{ii} = 1$, $d_i \in C[X]$ y no es difícil probar que $e_i = 0$. Además, si $d_i \in C[X]$ y la matriz X es de rango completo, existe una reparametrización del modelo tal que $X\beta = d_i\phi + a_1\gamma_1 + \dots + a_p\gamma_p$ lo que indica que hay un único parámetro dedicado a un único caso. Como señalan Cook y Weisberg (1982), tal situación es de rara aparición en la práctica.

4. LAS MATRICES Z_i^* ($i=1, 2, \dots, n$)

Los resultados del apartado 3 tienen una extensión natural a la detección simultánea de multicolinealidad y casos influyentes usando técnicas de análisis de condición en la matriz

$$Z_i^* = [Z^* \mid d_i]$$

de (3.4). Al efectuar un desarrollo paralelo al del apartado 3, se han de obtener, por fuerza, métodos de estudio simultáneo del "leverage ampliado" $\sigma_{ii}^*[Z]$ y de la condición de la matriz Z^* . Puesto que, según el capítulo II, esta última está relacionada con la condición de Z las técnicas de análisis de condición aplicadas sobre el conjunto de los Z_i^* ($i=1, \dots, n$) proporcionan información para la detección conjunta de casos influyentes y multicolinealidad.

4.1. La inversa de $Z_i^* Z_i^*$

Los resultados obtenidos de invertir $Z_i^* Z_i^*$ son interesantes y merecen ser estudiados en detalle.

Sea $i \in \{1, \dots, n\}$ fijo. La expresión de la matriz $Z_i^* Z_i^*$ es

$$Z_i^* Z_i^* = \begin{bmatrix} Z^{*'} \\ d_i^* \end{bmatrix} \begin{bmatrix} Z^* & | & d_i \end{bmatrix} = \left(\begin{array}{c|c} Z^{*'} Z^* & z_i^* \\ \hline z_i^{*'} & 1 \end{array} \right)$$

donde $z_i^* = (z_i^1, y_i)$ es la i -ésima fila de la matriz Z^* . Usando la técnica de inversión de una matriz en cajas, se tiene:

$$(Z_i^* Z_i^*)^{-1} = \left(\begin{array}{c|c} [Z^{*'} Z^* - z_i^* z_i^{*'}]^{-1} & -a_i^* \\ \hline -a_i^{*'} & \frac{1}{1 - \sigma_{ii}^* [z]} \end{array} \right) \quad (3.29)$$

donde $a_i^* = \frac{1}{1 - \sigma_{ii}^* [z]} (Z^{*'} Z^*)^{-1} z_i^*$. Interesan las expresiones explícitas del vector a_i^* y de la submatriz principal $m \times m$ en (3.29):

a) Expresión explícita de a_i^*

Según se obtuvo en (2.5), la inversa de $Z^{*'} Z^*$ es

$$(Z^{*'} Z^*)^{-1} = \left(\begin{array}{c|c} (Z'Z)^{-1} + \frac{\hat{\alpha} \hat{\alpha}'}{SCE} & -\frac{S_Y}{SCE} \hat{\alpha} \\ \hline -\frac{S_Y}{SCE} \hat{\alpha}' & \frac{S_Y^2}{SCE} = \frac{1}{1-R^2} \end{array} \right)$$

De aquí, el vector $m \times 1$ $(Z^{*'} Z^*)^{-1} z_i^*$ es

$$(Z^{*'} Z^*)^{-1} z_i^* = \left(\begin{array}{c} (Z'Z)^{-1} z_i + \frac{\hat{\alpha}}{SCE} [z_i' \hat{\alpha} - S_Y y_i] \\ \hline \frac{S_Y}{SCE} [S_Y y_i - z_i' \hat{\alpha}] \end{array} \right)$$

Finalmente teniendo en cuenta que

$$S_Y y_i - z_i' \hat{\alpha} = y_i - \bar{y} - z_i' \hat{\alpha} = e_i$$

la expresión explícita del vector a_i^* es

$$a_i^* = \left(\begin{array}{c} a_{p,i}^* \\ a_{m,i}^* \end{array} \right) = \frac{1}{1 - \sigma_{ii}^* [z]} \left(\begin{array}{c} (Z'Z)^{-1} z_i - \frac{e_i}{SCE} \hat{\alpha} \\ \hline \frac{e_i}{SCE} S_Y \end{array} \right) \quad (3.30)$$

Según (3.30)

$$a_{p,i}^* = \frac{1}{1 - v_{ii}^* [z]} \left[(Z'Z)^{-1} z_i - \frac{e_i}{SCE} \hat{\alpha} \right]$$

Haciendo las pertinentes operaciones no es difícil ver que

$$a_{p,i}^* = \frac{a_{m,i}^*}{S_y} \left[\frac{CIM_{z,i}}{(r_i^2/n-m)} - \hat{\alpha} \right] \quad (3.31)$$

donde, en (3.31), $CIM_{z,i} = (1 - v_{ii}^* [z])^{-1} (Z'Z)^{-1} z_i e_i$, en analogía con $CIM_i = (n-1) (\hat{\beta} - \hat{\beta}_{(i)}) = (n-1) \frac{(X'X)^{-1} x_i e_i}{1 - v_{ii}}$

Por otro lado, también de (3.30), es

$$a_{m,i}^* = \frac{1}{1 - v_{ii}^* [z]} \cdot \frac{e_i}{S_y SCE} = \quad (3.32)$$

$$= \frac{S_y}{(SCE)^{1/2}} \cdot \frac{1}{(1 - v_{ii}^* [z])^{1/2}} \cdot \frac{\frac{r_i [z]}{(n-m)^{1/2}}}{1 - \frac{r_i^2 [z]}{n-m}}$$

donde

$$r_i [z] = \frac{e_i}{\hat{\sigma} [1 - v_{ii}^* [z]]^{1/2}} \quad i = 1, 2, \dots, n \quad (3.33)$$

Obsérvese que

$$r_i [z] = \frac{e_i}{\hat{\sigma} [1 - v_{ii}^* [z]]^{1/2}} = \frac{e_i}{\hat{\sigma} [(1 - v_{ii}) + \frac{1}{n}]^{1/2}} \leq r_i$$

y, por tanto, $r_i^2 [z] \leq r_i^2 \leq n-m$. Además, si n es suficientemente grande, es $r_i [z] \approx r_i$ y, así, un "outlier" tiende a producir un $r_i [z]$ grande.

De (3.31), (3.32) y (3.33) un caso influyente, de por extremo $(v_{ii}^* [z], r_i^2 [z])$ ha de provocar valores altos en las coordenadas del vector $a_i^* = (a_{p,i}^*, a_{m,i}^*)'$.

b) Expresión explícita de $[Z^*Z^* - z_i^*z_i^{*'}]^{-1}$

Usando la técnica de inversión adecuada, la submatriz principal $m \times m$ de (3.29) es

$$[Z^*Z^* - z_i^*z_i^{*'}]^{-1} = (Z^*Z^*)^{-1} + \frac{(Z^*Z^*)^{-1} z_i^*z_i^{*'} (Z^*Z^*)^{-1}}{1 - v_{ii}^* [z]} = \quad (3.34)$$

$$= (Z^*Z^*)^{-1} + (1 - v_{ii}^* [z]) a_i^* a_i^{*'}$$

Así pues, la inversa (3.29) queda totalmente determinada con las fórmulas (3.30) y (3.34) junto con el resultado de inversión (2.5).

4.2. Diagnósticos de condición en Z_i^* ($i=1, \dots, n$)

4.2.1. Factores de inflación de la varianza

Para fijar la notación, se denotarán $VIF_{j,i}^*$ ($j=1, \dots, p$) los p primeros elementos diagonales de la matriz $(Z_i^* Z_i^*)^{-1}$, $VIF_{m,i}^*$ el m -simo y $VIF_{p+1,i}^*$ el $m+1$ y el último. Conviene dividir el análisis en tres partes diferenciadas.

a) Análisis de los $VIF_{j,i}^*$ ($j=1, \dots, p$)

Según (3.32) y (3.33), se tiene

$$VIF_{j,i}^* = VIF_j^* + [1 - v_{ii}^*[z]]^{-1} \left[c_j'(Z'Z)^{-1} z_i - \frac{e_i}{SCE} \hat{\alpha}_j \right]^2 \quad (3.35)$$

donde c_j es el j -ésimo vector canónico de \mathbb{R}^p . La expresión (3.35) se completa con la fórmula (2.7) para concluir

$$VIF_{j,i}^* = VIF_j^* + \frac{\hat{\alpha}_j^2}{SCE} + [1 - v_{ii}^*[z]]^{-1} \left[c_j'(Z'Z)^{-1} z_i - \frac{e_i}{SCE} \hat{\alpha}_j \right]^2 \quad (3.36)$$

$j = 1, \dots, p$

La fórmula (3.36) indica que valores altos de VIF_j^* provocan mala condición en la matriz Z_i^* reflejada en los $VIF_{j,i}^*$.

Del proceso de ortogonalización, es fácil deducir que $V_{2^*} = V_2 + \frac{ee'}{SCE}$ de donde

$$v_{ii}^*[z] = v_{ii}[z] + \frac{e_i^2}{SCE} \quad (3.37)$$

y de aquí

$$\begin{aligned} & [1 - v_{ii}^*[z]]^{-1} \left[c_j'(Z'Z)^{-1} z_i - \frac{e_i}{SCE} \hat{\alpha}_j \right]^2 = \\ & = \frac{\left(\frac{e_i}{SCE} \right)^2}{1 - v_{ii}[z] - \frac{e_i^2}{SCE}} \left[\frac{c_j'(Z'Z)^{-1} z_i}{\left(\frac{e_i}{SCE} \right)} - \hat{\alpha}_j \right]^2 = \\ & = \frac{1}{SCE} \cdot \frac{\frac{e_i^2}{SCE} [1 - v_{ii}[z]]}{1 - \frac{e_i^2}{SCE} [1 - v_{ii}[z]]} \left[\frac{c_j'(Z'Z)^{-1} z_i}{\left(\frac{e_i}{SCE} \right)} - \hat{\alpha}_j \right]^2 = (3.38) \\ & = \frac{1}{SCE} \cdot \frac{\frac{1}{n-m} r_i^2[z]}{1 - \frac{1}{n-m} r_i^2[z]} \left[\frac{c_j'(Z'Z)^{-1} z_i}{\left(\frac{e_i}{SCE} \right)} - \hat{\alpha}_j \right]^2 \end{aligned}$$

De (3.38) y (3.33), un "outlier", medido por $r_i^2[z]$, tiende a producir un $VIF_{m,i}^*$ alto. El corchete de (3.38) es de difícil interpretación.

b) $VIF_{m,i}^*$

La expresión de $VIF_{m,i}^*$ es

$$\begin{aligned} VIF_{m,i}^* &= VIF_m^* + \frac{S_y^2}{SCE} \cdot \frac{e_i^2 / SCE}{1 - \sigma_{ii}^*[z]} = \\ &= \frac{S_y^2}{SCE} + \frac{S_y^2}{SCE} \cdot \frac{e_i^2 / SCE}{1 - \sigma_{ii}^*[z]} = \quad (3.39) \\ &= VIF_m^* \left[1 + \frac{(1/n-m) r_i^2[z]}{1 - (1/n-m) r_i^2[z]} \right] \quad (\text{según (3.37)}) \end{aligned}$$

(3.39) demuestra que $VIF_{m,i}^*$ es grande si el caso i -ésimo es un "outlier" o si VIF_m^* es grande (o ambas cosas a la vez).

c) VIF^{*i}

Se deduce de (3.29)

$$VIF^{*i} = \frac{1}{1 - \sigma_{ii}^*[z]} \quad (3.40)$$

expresión que informa directamente de la influencia del caso i -ésimo medida por $\sigma_{ii}^*[z]$.

Un interés adicional de los factores de inflación de la varianza en la matriz ampliada es consecuencia del siguiente resultado:

LEMA 3.3.

Sea A una matriz $S \times S$ definida positiva. Si $A = (a_{jk})$ $j=1, \dots, S$ $k=1, \dots, S$ es

$$|a_{jk}| \leq (a_{jj})^{1/2} (a_{kk})^{1/2} \quad j=1, \dots, S \quad k=1, \dots, S$$

DEMOSTRACION

Basta observar que $a_{jk} = c_j' A c_k$, donde c_j y c_k son los vectores canónicos correspondientes, y aplicar la desigualdad de Cauchy-Schwartz.

Tomando en el lema $A = (Z_i^* Z_i^*)^{-1}$, es

$$|a_{jp,i}^*| \leq (VIF_{j,i}^*)^{1/2} (VIF^*)^{1/2} \leq \max_{j=1,\dots,m} \{ \max_{i=1,\dots,p} VIF_{j,i}^*, VIF^* \}$$

$$y \quad |a_{m,i}^*| \leq (VIF_{m,i}^*)^{1/2} (VIF^*)^{1/2} \leq \max_{j=1,\dots,m} \{ \max_{i=1,\dots,p} VIF_{j,i}^*, VIF^* \} \quad (3.41)$$

donde $a_{p,i}^* = (a_{ip}^*, \dots, a_{pp}^*)'$. De (3.41) puede concluirse que un caso influyente, de valor anómalo en alguna de las coordenadas $(\sigma_{ii}[z], r_i^2[z])$ tiende a provocar mala condición en la matriz Z_i^* al provocar valores altos en las coordenadas de a_i^* según se deduce de (3.31) y (3.32).

Finalmente, es válido el análogo del teorema 3.1 que se enuncia sin demostración.

TEOREMA 3.6.

Sea la matriz de $(m+1) \times (m+1)$

$$\begin{pmatrix} (Z^* Z^*)^{-1} & 0 \\ 0 & 1 \end{pmatrix} \quad (3.42)$$

Se cumple que la norma euclídea de la diferencia entre $(Z_i^* Z_i^*)^{-1}$ y (3.42) tiende a cero cuando $\sigma_{ii}[z] \rightarrow 0$.

4.2.2. Otros resultados

4.2.2.a) Autovalores e índices de condición

Sean $0 < \lambda_{1,i}^* \leq \lambda_{2,i}^* \leq \dots \leq \lambda_{m,i}^* \leq \lambda_{m+1,i}^*$ los autovalores de la matriz $Z_i^* Z_i^*$. Usando técnicas de demostración ya conocidas, puede establecerse el teorema de separación entre estos autovalores y los de $Z^* Z^*$.

TEOREMA 3.7.

Si $k_i^* = C^* z_i^*$ es un vector $m \times 1$ cuyas coordenadas son todas distintas de cero y los autovalores de $Z^* Z^*$ son todos distintos entre sí, entonces con probabilidad uno se cumple

$$0 < \lambda_{1,i}^* < \lambda_1^* < \lambda_{2,i}^* < \dots < \lambda_m^* < \lambda_{m+1,i}^* \quad (3.43)$$

De (3.43) pueden deducirse las relaciones correspondientes entre los índices de condición de la matriz Z_i^* . Sean

$$\eta_k^* = \left[\frac{\lambda_{\max}^*}{\lambda_k^*} \right]^{1/2}, \quad \eta_{k,i}^* = \left[\frac{\lambda_{\max,i}^*}{\lambda_{k,i}^*} \right]^{1/2}$$

$$(k=1, \dots, p, m) \quad (k=1, \dots, p, m, m+1)$$

y

$$\delta_{k,i}^* = \left[\frac{\lambda_{m,i}^*}{\lambda_{k+1,i}^*} \right]^{1/2} \quad (k=1, \dots, m-1 (=p))$$

Usando el teorema de separación 3.7 es inmediato que

$$\delta_{k,i}^* < \eta_k^* < \eta_{k,i}^*, \quad (k=1, \dots, m-1 (=p)) \quad (3.44)$$

Tomando $k=1$ en (3.44) se obtiene para el número de condición

$$\delta_{1,i}^* < k[Z^*] < k[Z_i^*] \quad (3.45)$$

En analogía con el teorema 3.3 puede establecerse el siguiente resultado

TEOREMA 3.8.

Se tiene que $[\lambda_{1,i}^*, \dots, \lambda_{m+1,i}^*] \rightarrow [\lambda_1^*, \dots, \lambda_{m,i}^*]$ cuando $\sigma_{ii}^*[Z] \rightarrow 0$.

4.2.2.b) Módulo de Gunst

Al proceder como en el apartado (3.1.3) se llega a la identidad

$$\lambda_{1,i}^* = \|Z_i^* \gamma_{\min,i}^*\|^2 = \|Z^* [\gamma_{1,i}^* + \gamma_{m,i}^* c_i^*]\|^2 + \gamma_{m,i}^{*2} [1 - \sigma_{ii}^*[Z]] \quad (3.46)$$

donde $\gamma_{\min,i}^* = [\gamma_{1,i}^*, \gamma_{m,i}^*]^T$ es un autovector normalizado asociado al menor autovalor de $Z_i^* Z_i^*$ y c_i^* es tal que $V_2^* d_i = Z^* c_i^*$ siendo V_2^* la matriz de proyección sobre la variedad $C[Z^*]$. Se cumple

TEOREMA 3.9.

Si $Z_i^* \neq 0$, se verifica

$$a) \quad \|\gamma_{1,i}^* + \gamma_{m,i}^* c_i^*\| \neq 0$$

b) La desigualdad

$$0 < \lambda_{1,i}^* < \lambda_1^* \leq \left\| \frac{Z^* [\gamma_{1,i}^* + \gamma_{m,i}^* c_i^*]}{\|\gamma_{1,i}^* + \gamma_{m,i}^* c_i^*\|} \right\|^2 =$$

$$= \frac{\lambda_{1,i}^* - \gamma_{m,i}^{*2} [1 - \sigma_{ii}^*[Z]]}{\|\gamma_{1,i}^* + \gamma_{m,i}^* c_i^*\|^2} \quad (3.47)$$

DEMOSTRACION:

- a) Análoga a la del lema 3.1.
- b) Análoga a la del teorema 3.4. #

4.2.2.c) Desigualdades de Berk

No es difícil probar, basándose en la técnica de demostración del teorema de Berk que se tiene el

TEOREMA 3.10

$$\begin{aligned} \max \{ VIF_{1,i}^*, \dots, VIF_{m,i}^*, VIF_i^{*i} \} &\leq k^2 [Z_i^*] \leq \\ &\leq (m+1) \left[VIF_i^{*i} + \sum_{j=1}^m VIF_{j,i}^* \right] \end{aligned} \quad (3.48)$$

Una primera consecuencia inmediata de (3.48) es que $[1 - VIF_i^{*i} [Z]]^{-1} \leq k^2 [Z_i^*]$ de modo que casos influyentes tienden a provocar mala condición en Z_i^* .

Según la desigualdad de Cauchy-Schwartz es $VIF_{j,i}^* \leq VIF_i^{*i} \cdot VIF_j^*$ ($j=1, \dots, m$) directamente de (3.34) y, por consiguiente,

$$k^2 [Z_i^*] \leq (m+1) VIF_i^{*i} \left[1 + \sum_{j=1}^m VIF_j^* \right] \quad (3.49)$$

Del teorema 2.2

$$\sum_{j=1}^m VIF_j^* \leq m \max_j VIF_j^* \leq m k^2 [Z^*]$$

de manera que, continuando en (3.49), se tiene

$$k^2 [Z_i^*] \leq (m+1) VIF_i^{*i} [1 + m k^2 [Z^*]] \quad (3.50)$$

Uniendo (3.45) y (3.50) se llega al siguiente resultado

LEMA 3.3.

$$k^2 [Z^*] < k^2 [Z_i^*] \leq (m+1) VIF_i^{*i} [1 + m k^2 [Z^*]] \quad (3.51)$$

4.3. Detección de multicolinealidad y casos influyentes

El método que se deduce del apartado 4.2. de detección simultánea de multicolinealidad y casos influyentes puede expresarse de acuerdo con las siguientes etapas:

- i) Estudiar, en primer lugar, la columna VIF_i^{*i} ($i=1, \dots, n$)

(mediante, p. ejemplo, un "index plot"). Puesto que de (3.37) se tiene $\sigma_{ii}^*[z] = 1 - [1 - \sigma_{ii}[z]] \left[1 - \frac{r_i^2[z]}{n-m} \right]$ según (3.40), VIF^{*i} crece con las coordenadas del bloque $(\sigma_{ii}[z], r_i^2[z])$. Por consiguiente, los casos influyentes se destacarán como puntos anómalos en un "index plot" de los VIF^{*i} .

ii) Como complemento al "index plot" de los VIF^{*i} , el siguiente paso es distinguir si la influencia del caso $z_i^* = (z_i', y_i)$ se debe a que z_i es extremo o a que y_i es "outlier". Para ello, obsérvese, en primer lugar, que de (3.38) se tiene

$$\frac{VIF_{m,i}^*}{VIF_m^*} = \left[1 - \frac{1}{n-m} r_i^2 \right]^{-1} \quad (3.52)$$

y en segundo lugar, de (3.40) y (3.37),

$$\begin{aligned} VIF^{*i} &= \frac{1}{1 - \sigma_{ii}^*[z]} = \frac{1}{1 - \sigma_{ii}[z] - \frac{e_i^2}{SCE}} = \\ &= \frac{1/(1 - \sigma_{ii}[z])}{1 - (1/(n-m)) r_i^2[z]} = VIF^i \frac{VIF_{m,i}^*}{VIF_m^*} \end{aligned} \quad (3.53)$$

Según (3.53),

$$VIF^i = \frac{VIF_m^* \cdot VIF_{m,i}^*}{VIF_{m,i}^*} \quad (3.54)$$

Las expresiones (3.52) y (3.53) dependen de información deducida de la inversa (3.29) y del estadístico VIF_m^* que, según la fórmula (2.8), puede deducirse de la tabla de análisis de la varianza del modelo. Un "index plot" de (3.51) detecta las respuestas anómalas mientras que el "index plot" análogo de VIF^i detecta los puntos extremos.

iii) Sea i_0 tal que $VIF^{*i_0} = \min VIF^{*i}$. En virtud de los resultados del apartado 4.2, si $\sigma_{i_0 i_0}^*[z] \approx 0$, los diagnósticos de multicolinealidad de $Z_{i_0}' Z_{i_0}^*$ están próximas a los de $Z^* Z$. Las técnicas del Capítulo II, pueden utilizarse para detectar la multicolinealidad en Z .

4.4. Ejemplo

La metodología completa basada en las matrices Z_i^* , ($i = 1, \dots, n$) puede aplicarse al ejemplo tratado en 3.6.

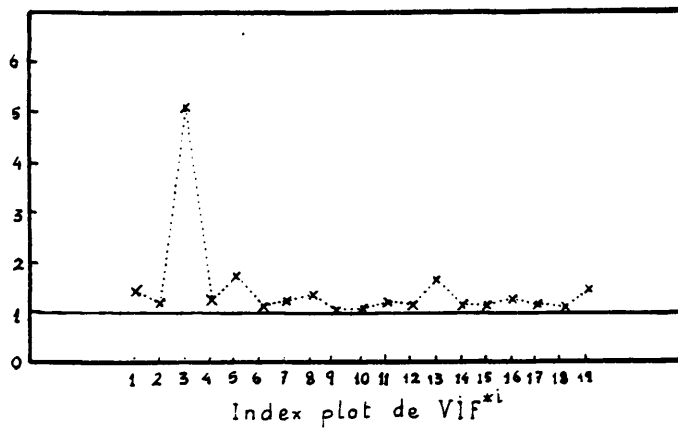


Fig. 3.2

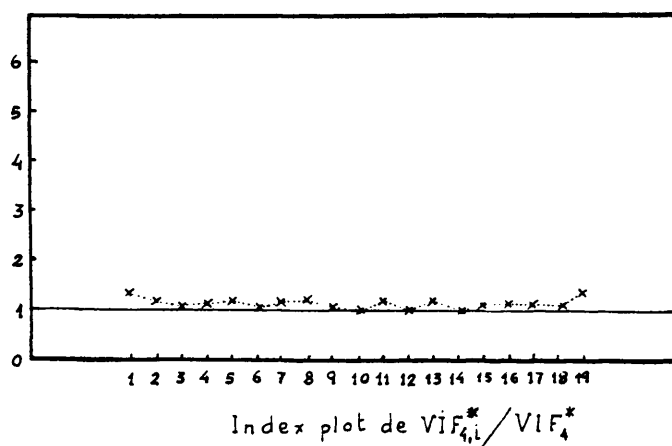


Fig 3.3

La figura 3.2 es un "index plot" de los VIF^{*i} ($i=1, \dots, n$). Se detecta inmediatamente el caso 3 como punto destacado. La figura 3.3. no indica nada anormal por lo que, volviendo a la figura 3.1, la peculiaridad del caso 3 es que x_3 está alejado del centro de los datos.

De nuevo, $\min_i VIF^{*i} = VIF^{*9} = 1.016$ ($\sigma_{99}^2 = 0.016$).

La tabla 3.4.a) contiene los valores de $VIF_{j,q}^*$ ($j=1,2,3$) y $k^2[Z_q^*]$. Por su parte, la tabla 3.4.b) contiene los pasos necesarios para el cálculo de la cota (3.47)

i_o	$VIF_{1,q}^*$	$VIF_{2,q}^*$	$VIF_{3,q}^*$	$k^2[Z_q^*]$
9	75.8156	1.3962	76.1979	368.79

Tabla 3.4.a)

i_o	$\lambda_{1,q}$	$r_{1,q}^{*2} [1-v_{1,q}^2(\omega)]$	$\ r_{1,q}^* + r_{2,q}^* c\ ^2$
9	0.0066	0.0000814	0.9755

Tabla 3.4.b)

Como $\sigma_{q,q}^*[z] \approx 0$, el teorema 3.6 permite asegurar que $VIF_{j,q}^* \approx VIF_j^*$ ($j=1,2,3$) con lo que los resultados de 3.4.a) identifican una multicolinealidad muy severa en Z^* . Por otro lado, el método de la cota (2.28) permite asegurar que

$$\lambda_1^* \in (0.0066, 0.0067)$$

con lo cual $\lambda_1^* \approx 0.0066$. Finalmente, el teorema 3.8 garantiza que $k^2[Z_q^*] \approx k^2[Z^*]$ de manera que $k^2[Z^*] \approx 368.79$

Una vez detectada la multicolinealidad en Z^* , el siguiente paso es estudiar la multicolinealidad en Z usando las técnicas del capítulo II. En este ejemplo, no es necesario calcular los diagnósticos de Z^* puesto que al ser $\sigma_{q,q}^*[z] \approx 0$ puede garantizarse que $(Z_q^* Z_q^*)^{-1} \approx (Z^* Z^*)^{-1}$ y que el autosistema de $Z_q^* Z_q^*$ está próximo al de $(Z^* Z^*)^{-1}$. La tabla 3.5.a) recoge la aproximación a la inversa de $Z^* Z^*$ y la 3.5.b) la aproximación a λ_1^* y γ_{\min}^* y los demás autovalores de $Z^* Z^*$.

75.816	-2.2098	-75.3309	6.1620
.	1.3962	1.5735	-0.3084
.	.	76.1979	-6.3044
.	.	.	1.5688

TABLA 3.5.a): Aproximación de $(Z^* Z^*)^{-1}$

λ^*	1	r_{min}^*			
0.0066	1	-0.7349	0.0099	0.7066	-0.00586
$\lambda_2^* \approx 0.6275$		$\lambda_3^* \approx 0.9208$		$\lambda_4^* \approx 2.4340$	

Aproximación al autosistema Z^*Z^*

Tabla 3.5.b)

A partir de aquí se procede como en el ejemplo del Capítulo II. Siguiendo el criterio de (2.11),

$$\frac{(z^{*14})^2}{z^{*44}} = 24.203 \approx 24.816 = 75.816 - 51 = z^{*14} - 51$$

$$\frac{(z^{*24})^2}{z^{*44}} = 0.061 > -49.604 = 1.3962 - 51 = z^{*22} - 51$$

$$\frac{(z^{*34})^2}{z^{*44}} = 25.331 \approx 25.998 = 76.198 - 51 = z^{*33} - 51$$

se deduce que $VIF_1 \approx 51$ y $VIF_3 \approx 51$ lo que identifica serios problemas de multicolinealidad relacionados con las variables x_1 y x_3 .

El cálculo de la desigualdad (2.31) informa que

$$\lambda_1 \in (0.0066, 0.0099)$$

lo que implica que λ_1 es pequeño (de hecho, según la tabla 3.3 es $\lambda_1 = 0.0099$). Obsérvese que, en este problema, al contrario de lo que ocurría en el capítulo II, la multicolinealidad asociada a λ_1^* es no predictiva y la desigualdad (2.31) resulta de utilidad. Finalmente, el vector de multicolinealidad

$$\frac{Y_1^* + \hat{\alpha}/s_y Y_m^*}{\|Y_1^* + \hat{\alpha}/s_y Y_m^*\|} \text{ es } (-0.7180; 0.0095; 0.7044)' \text{ que identifica}$$

una dependencia lineal aproximada entre x_1 y x_3 . Las cotas que se obtienen como consecuencia del teorema de separación 2.1 son

$$1.211 < k[Z] < 19.204$$

$$1 < \eta_2 < 1.369$$

que tampoco aquí son de especial interés.

Las tablas 3.6.a) y 3.6.b) contienen la inversa de Z^*Z^* y su autosistema. Es de destacar la proximidad a las tablas

3.5.a) y 3.5.b), respectivamente.

74.8456	-2.2076	-74.4117	6.1111
.	1.3961	1.5657	-0.3080
.	.	75.3267	-6.2562
.	.	.	1.5661

TABLA 3.6.a) Inversa de Z^*Z^*

j	λ_j^*	Y_j^*			
1	0.0067	-0.7046	0.0180	0.7069	-0.0588
2	0.6275	0.2800	-0.8888	0.3165	0.1771
3	0.9365	-0.2388	0.0583	-0.1600	0.9560
4	2.4293	-0.6067	-0.4542	-0.6120	-0.2262

TABLA 3.6.b) Autosistema de Z^*Z^*

4.5. Datos sin tipificar

No hay modificaciones sustanciales en las expresiones formales de los resultados obtenidos sin tipificar los datos considerando la matriz

$$X_i^* = [X^* | d_i]$$

La inversa de $X_i^{*'} X_i^*$ es

$$(X_i^{*'} X_i^*)^{-1} = \left(\begin{array}{c|c} [X_{(i)}^{*'} X_{(i)}^*]^{-1} & -a_i^* \\ \hline -a_i^{*'} & \frac{1}{1 - \sigma_{ii}^*} \end{array} \right)$$

donde $a_i^* = [1 - \sigma_{ii}^*]^{-1} (X_i^{*'} X_i^*)^{-1} x_i^*$

de modo que, en notación obvia, se tiene

$$VIF^{*i} = \frac{1}{1 - \sigma_{ii}^*}$$

y

$$VIF_{j,i}^* = VIF_{j(i)}^* \quad j = 1, 2, \dots, m, m+1 \quad (3.55)$$

Es interesante reconocer la forma explícita de (3.55) para una posterior utilización. No es difícil ver que

$$VIF_j^*(i) = VIF_{j,i}^* = VIF_j^* + (1 - \sigma_{ii}^*) a_i^* a_i^{*'} \quad (3.56)$$

$j = 1, \dots, m, m+1$

donde ahora la expresión de a_i^* es

$$a_i^* = \frac{1}{1 - \sigma_{ii}^*} \left(\frac{(X'X)^{-1} x_i - \frac{e_i}{SCE} \hat{\beta}}{\frac{e_i}{SCE}} \right)$$

Por consiguiente

$$\begin{aligned} VIF_j^*(i) &= VIF_j^* + \frac{1}{1 - \sigma_{ii}^*} \left[c_j^* (X'X)^{-1} x_i - \frac{e_i}{SCE} \hat{\alpha}_j \right]^2 = \\ &= VIF_j^* + \frac{1}{SCE} \cdot \frac{\frac{n-m}{r_i^2}}{1 - \frac{r_i^2}{n-m}} \left[\frac{DFBETA_{ij}}{\left(\frac{r_i^2}{n-m} \right)} - \hat{\beta}_j \right]^2 \end{aligned} \quad (3.57)$$

$j = 1, \dots, m$

lo que demuestra que la distancia $|VIF_{j,i}^* - VIF_j^*|$ es igual a una expresión que depende de elementos que determinan la influencia del caso i -ésimo. Por otro lado, se tiene

$$\begin{aligned} VIF_{m+1,i}^* &= VIF_{m+1}^*(i) = \frac{1}{SCE} + \frac{1}{1 - \sigma_{ii}^*} \cdot \frac{e_i^2}{SCE} = \\ &= VIF_{m+1}^* + \frac{\frac{r_i^2}{n-m}}{1 - \frac{r_i^2}{n-m}} \end{aligned}$$

por lo que

$$|VIF_{m+1}^* - VIF_{m+1}^*(i)| = \frac{\frac{r_i^2}{n-m}}{1 - \frac{r_i^2}{n-m}} \quad (3.58)$$

es función monótona del estadístico r_i^2 .

CAPITULO IV

1. SUMARIO

En este capítulo, se desarrolla un estudio de influencia basado fundamentalmente en ciertos diagnósticos de multicolinealidad, bien de la matriz de diseño o bien de la matriz ampliada.

La sección 2 es una presentación del problema, de los objetivos que se pretenden y de las dificultades matemáticas existentes. Ante la carencia de fórmulas explícitas para evaluar, de modo general, la influencia de un caso individual en los diagnósticos de multicolinealidad, es conveniente la elaboración de una técnica que permita paliar esa deficiencia. La sección 3 recoge un método aproximado o asintótico de análisis basado en propiedades de diferenciabilidad de funcionales que dependen de la función de distribución empírica de los n casos del modelo.

Por su parte, la sección 4 está dedicada al autosistema de la matriz de productos cruzados. La aplicación de la técnica de la sección 3 a los autovalores de dicha matriz, conduce a resultados importantes. Es interesante destacar que la utilización de este método permite, en el apartado 4.2.2.b) una nueva interpretación del comportamiento de los gráficos de componentes principales. Los resultados obtenidos considerando los autovalores como funcionales se complementan con otros deducidos de técnicas clásicas de la teoría de perturbaciones. Por lo que se refiere a los autovectores, se analiza, en un caso particular, la influencia de un caso mediante el criterio de Wilks de detección de "outliers".

La sección 5 se dedica a los factores de inflación de la varianza. El apartado 5.1 estudia la influencia de un caso mediante la aplicación de fórmulas explícitas, relacionando este estudio con la distancia de Cook. En el 5.2 se utiliza el método aproximado de la sección 3.

Para finalizar, la sección 6 se compone de un estudio de influencia en otros diagnósticos tales como los elementos y determinante de la matriz de correlaciones.

Se incluyen ejemplos prácticos en los apartados 3.4 y 4.2.

2. MOTIVACION

Supóngase que bajo el concepto de diagnóstico de multicolinealidad se agrupan diagnósticos calculados bien para la matriz $X \text{ o } X^*$, ya tipificando los datos o no. Como ha quedado señalado en la introducción, la idea general de un análisis de influencia es estudiar los cambios provocados en ciertos aspectos del análisis de un modelo lineal al aplicar un determinado esquema de perturbación a los datos. Por consiguiente, puede admitirse legítimamente dentro del problema de influencia el estudio del cambio producido en los diagnósticos de multicolinealidad al perturbar la matriz de diseño y/o el vector de observaciones. Precizando más, si la perturbación que se considera obedece al esquema de eliminación de un único caso, en este capítulo se trata de resolver problemas como los que siguen: si D^* es un diagnóstico basado en X^* y $D^*_{(i)}$ es el correspondiente diagnóstico basado en $X^*_{(i)} = [X^*_{(i)} | Y^*_{(i)}]$, ¿el análisis de la relación entre D^* y $D^*_{(i)}$ ¿qué información produce? ¿qué características de un caso intervienen en el cambio de D^* a $D^*_{(i)}$? ¿está relacionado el cambio de D^* a $D^*_{(i)}$ con la influencia de un caso? Si la respuesta a esta última pregunta es afirmativa, ¿en qué sentido?.

El problema planteado en el primer párrafo es una reformulación de la idea básica de la que parten Andrews y Pregibon (1978) en la construcción del estadístico R_i :

$$R_i = \frac{|X^{*'}_{(i)} X^*_{(i)}|}{|X^{*'} X^*|} = 1 - v_{ii}^* \quad (4.1)$$

Teniendo en cuenta que el determinante de la matriz $X^{*'} X^*$ es un indicador de la estabilidad numérica de la inversa, el estadístico R_i no es más que la razón entre el determinante perturbado $|X^{*'}_{(i)} X^*_{(i)}|$ y el determinante original $|X^{*'} X^*|$.

La idea de este capítulo está, pues, ya implícita en la consideración de Andrews-Pregibon. En este capítulo, se emplea sistemáticamente este enfoque en relación con diagnósticos de multicolinealidad más usuales tales como el autosistema de la matriz de productos cruzados y los factores VIF_j .

Devlin et al. (1975), estudiando el cambio en el coeficiente de correlación muestral $\hat{\rho}$ asociado a una m.a.s. de una población bidimensional demuestran la fórmula

$$\hat{\rho}_{(i)} = \frac{\hat{\rho} - \frac{r_{1i} r_{2i}}{n}}{\left(1 - \frac{r_{1i}^2}{n}\right)^{1/2} \left(1 - \frac{r_{2i}^2}{n}\right)^{1/2}} \quad (4.2)$$

donde $r_{ji} = (x_{ji} - \bar{x}_j) / \left[\tilde{S}_j^2 (1 - 1/n) \right]^{1/2}$, $\tilde{S}_j^2 = \frac{1}{n} \sum (x_{ji} - \bar{x}_j)^2$ ($j=1,2$).

Al contrario de lo que ocurre en (4.1) y (4.2), no siempre existe una expresión explícita sencilla para analizar el cambio en el diagnóstico al eliminar un caso. Esto se hace particularmente notable en el estudio del autosistema de la matriz de productos cruzados. Por consiguiente, una posible solución es desarrollar una técnica aproximada, válida en condiciones generales, para analizar la relación entre diagnósticos genéricos D^* y $D^*_{(i)}$.

3. DIAGNOSTICOS COMO FUNCIONALES

Sea F_n^* la función de distribución empírica asociada a los n casos $x_i^* = (x_i^1, y_i^1)$ ($i=1, \dots, n$) de la matriz X^* . Los diagnósticos de multicolinealidad en X^* pueden escribirse en la forma

$$D^* = k_n T[F_n^*] \quad (4.3)$$

donde k_n es una constante correctora y $T[F_n^*]$ es un funcional adecuado. En efecto, no hay más que observar que

$$\frac{X^{*'} X^*}{n} = \int_{\mathbb{R}^m} \begin{pmatrix} x \\ y \end{pmatrix} (x^{*'} y) dF_n^* = A[F_n^*] \quad (4.4)$$

es un funcional en F_n^* . Por ejemplo, si $D^* = |X^{*'} X^*|$ es

$$D^* = n^m \left| \frac{X^{*'} X^*}{n} \right| = n^m |A[F_n^*]| \quad \text{que es de la forma (4.3) con}$$

$$k_n = n^m \quad T[F_n^*] = |A[F_n^*]|. \text{ Según (4.3), se tiene}$$

$$D^* - D^*_{(i)} = k_n T[F_n^*] - k_{n-1} T[F_{n-1}^*(i)] \quad (4.5)$$

y, de (4.5), el cambio $D^* - D^*_{(i)}$ depende de la relación entre $T[F_n^*]$ y $T[F_{n-1}^*(i)]$. De hecho, este apartado resuelve el problema de la aproximación de $T[F_n^*] - T[F_{n-1}^*(i)]$. Una vez resuelto este punto, la fórmula (4.5) permite hallar la aproximación adecuada en cada caso particular.

Es importante resaltar que cuando el estudio se base en los diagnósticos asociados a la matriz tipificada Z^* todos estos tendrán una expresión similar a (4.3) con $k_n = 1$. Basta demostrar que $Z^{*'} Z^* = B[F_n^*]$. En efecto, si X^* es la matriz de datos centrados asociada a X^* es

$$\begin{aligned} \frac{1}{n} X^{*'} X^* &= \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} x_i \\ y_i \end{pmatrix} (x_i' \ y_i') - \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} (\bar{x}' \ \bar{y}') = \\ &= \int_{\mathbb{R}^m} \begin{pmatrix} x \\ y \end{pmatrix} (x' \ y') dF_n^* - \left[\int_{\mathbb{R}^m} \begin{pmatrix} x \\ y \end{pmatrix} dF_n^* \right] \left[\int_{\mathbb{R}^m} \begin{pmatrix} x \\ y \end{pmatrix} dF_n^* \right]' \\ &= C[F_n^*] \end{aligned} \quad (4.6)$$

Si \otimes representa el producto de Hadamard entre dos matrices (Rao, 1973, pág. 30), entonces, según (4.6)

$$\begin{aligned} Z^{*'} Z^* &= \text{diag} \left(\frac{1}{S_1}, \dots, \frac{1}{S_p}, \frac{1}{S_y} \right) X^{*'} X^* \text{diag} \left(\frac{1}{S_1}, \dots, \frac{1}{S_p}, \frac{1}{S_y} \right) = \\ &= [I_m \otimes C[F_n^*]]^{1/2} C[F_n^*] [I_m \otimes C[F_n^*]]^{1/2} = \\ &= B[F_n^*] \end{aligned} \quad (4.7)$$

donde, en (4.7), $S_j^2 = \sum_{i=1}^n (x_{ij} - \bar{x}_{.j})^2 \quad (j = 1, \dots, p)$.

Evidentemente el estudio que se hace para $D^* = T[F_n^*]$ puede desarrollarse paralelamente para la expresión $D = T[F_n]$, donde D es el diagnóstico basado en la matriz X y F_n es la función de distribución de la medida de diseño asociada a las n filas de X , sin más que efectuar leves retoques de carácter formal.

La relación entre F_n^* y F_{n-1}^* es tal que permite hallar una relación aproximada entre D^* y $D^*_{(i)}$ explotando ciertas propiedades de diferenciabilidad del funcional $T[F_n^*]$.

3.1. Diferenciabilidad de $D^* = T[F_n^*]$

Para obtener la diferenciabilidad Fréchet de todos los diagnósticos $D^* = T[F_n^*]$ considerados como funcionales en F_n^* basta una aplicación apropiada en forma secuencial de los resultados del apéndice A.2.3. junto con la regla de la cadena de A.2.1. Todas estas técnicas están recogidas en el apéndice A.2.

3.2. Fórmula para $D^* - D^*(i)$

Una vez obtenida la diferenciabilidad Fréchet de $D^* = T[F_n^*]$ es fácil establecer la aproximación anunciada para $D^* - D^*(i)$.

Si F_n^* es la f. de distribución empírica calculada con toda la muestra, y $F_{n-1}^*(i)$, la f. de distribución empírica asociada a la muestra de tamaño $n-1$ que resulta de eliminar el caso i -ésimo, es fácil comprobar que

$$F_{n-1}^*(i) = F_n^* + \frac{1}{n-1} [F_n^* - \delta(x_i^*)] \quad (4.8)$$

y

$$F_n^* = F_{n-1}^*(i) + \frac{1}{n} [\delta(x_i^*) - F_{n-1}^*(i)] \quad (4.9)$$

donde $\delta(x_i^*)$ es la f. de distribución causal en el caso $x_i^* = (x_i^1, y_i)$

Según (4.8), si $D^* = T[F_n^*]$, se tiene:

$$\begin{aligned} -[D^* - D^*(i)] &= T[F_{n-1}^*(i)] - T[F_n^*] = \\ &= T\left[F_n^* + \frac{1}{n-1} (F_n^* - \delta x_i^*)\right] - T[F_n^*] = \\ &= DT\left[F_n^*, \frac{1}{n-1} (F_n^* - \delta x_i^*)\right] + o[\|F_n^* - F_{n-1}^*(i)\|_\sigma] \end{aligned} \quad (4.10)$$

donde

$$\begin{aligned} DT\left[F_n^*, \frac{1}{n-1} (F_n^* - \delta x_i^*)\right] &= \\ &= \lim_{t \rightarrow 0} \frac{T\left[F_n^* + \frac{t}{n-1} (F_n^* - \delta x_i^*)\right] - T[F_n^*]}{t} = \\ &= -\left(\frac{1}{n-1}\right) \lim_{t^* \rightarrow 0} \frac{T[F_n^* + t^*(\delta x_i^* - F_n^*)] - T[F_n^*]}{t^*} = -\frac{CIE_i}{n-1} \end{aligned} \quad (4.11)$$

límite que existe en virtud de la definición.

En (4.11), se ha adoptado la notación CIE_i = (curva de influencia empírica) por la semejanza del límite del miembro izquierdo con la curva de influencia verdadera de un estimador considerado como un funcional en F_n^* (Huber (1981), Hampel, (1974), Cuevas (1980)). Sustituyendo, finalmente, (4.11) en (4.10) se sigue, finalmente, que

$$D^* - D^*(i) = \frac{1}{n-1} CIE_i + o[\|F_n^* - F_{n-1}^*(i)\|_\sigma] \quad (4.12)$$

Partiendo de (4.9) y siguiendo un camino parecido puede establecerse un análogo de (4.12)

$$D^* - D_{(i)}^* = \frac{1}{n} CIE_{(i)}[x_i^*] + o\left[\|F_n^* - F_{n-1(i)}^*\|_{\sigma}\right] \quad (4.13)$$

donde $CIE_{(i)}[x_i^*]$ es la i -ésima curva de influencia empírica con el caso i -ésimo eliminado. (C-Weisberg, 1982).

Las expresiones (4.12) y (4.13) son la base de la aproximación para el análisis de $D^* - D_{(i)}^*$. En cada diagnóstico particular, basta calcular la correspondiente CIE_i (o $CIE_{(i)}[x_i^*]$). Desde el punto de vista práctico, la más adecuada es (4.12) porque solo necesita de los valores de la CIE_i para $i=1, 2, \dots, n$ mientras que, por el contrario, (4.13) requiere del conocimiento de las n curvas $CIE_{(i)}[x_i^*]$.

3.3. Calidad y mejora de la aproximación

En virtud de la fórmula (4.8)

$$\|F_n^* - F_{n-1(i)}^*\|_{\sigma} = \frac{1}{n-1} \|F_n^* - \delta x_i^*\|_{\sigma} \leq \frac{2}{n-1}$$

de manera que $(n-1)\|F_n^* - F_{n-1(i)}^*\|_{\sigma} = O(1)$ y, por tanto, siguiendo en (4.12)

$$\frac{o[\|F_n^* - F_{n-1(i)}^*\|_{\sigma}]}{(1/n-1)} = \frac{o[\|F_n^* - F_{n-1(i)}^*\|_{\sigma}]}{\|F_n^* - F_{n-1(i)}^*\|_{\sigma}} (n-1)\|F_n^* - F_{n-1(i)}^*\|_{\sigma} \rightarrow O(n \rightarrow \infty)$$

Así, el resto en (4.12) y (4.13) es $o\left(\frac{1}{n}\right)$ y, por consiguiente, el desarrollo de 3.2 se puede completar de la forma

$$D^* - D_{(i)}^* = \frac{1}{n-1} CIE_i + o\left(\frac{1}{n}\right) \quad (4.14)$$

y

$$D^* - D_{(i)}^* = \frac{1}{n} CIE_{(i)}[x_i^*] + o\left(\frac{1}{n}\right)$$

La mejora de la aproximación depende de distintos factores. Si, por ejemplo, $n^c \|F_n^* - F_{n-1(i)}^*\|_{\sigma} = O(1)$, entonces $D^* - D_{(i)}^* = \frac{1}{n-1} CIE_i + o\left(\frac{1}{n^{c+1}}\right)$. Además, los resultados teóricos del apartado A.2 permiten establecer, además de la simple diferenciabilidad Fréchet, la holomorfía de $D^* = T[F_n^*]$ como argumento de la f. de distribución empírica de la muestra x_1^*, \dots, x_n^* .

Si, por ejemplo, se desarrolla D^* hasta la segunda derivada, se tendrá

$$- [D^* - D^*_{(i)}] = T[F^*_{n-1(i)}] - T[F^*_n] = DT\left[F^*_n, \frac{1}{n-1}(F^*_n - \delta x^*_i)\right] + \\ + DT^{(2)}\left[F^*_n, \frac{1}{n-1}(F^*_n - \delta x^*_i)\right] + o\left(\frac{1}{n^2}\right)$$

Este último método podrá ser de utilidad si la expresión de la aplicación bilineal $DT^{(2)}$ no es excesivamente complicada.

3.4. Ejemplo

Aunque la teoría general para el autosistema se desarrolla en el apartado 4 de este mismo capítulo una buena forma de ilustrar la aplicación de esta técnica puede ser el siguiente ejemplo sencillo. Considérese un modelo de regresión lineal simple

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

y el problema del estudio de influencia en los autovalores de la matriz de correlación ampliada

$$Z^* Z^* = \begin{pmatrix} 1 & \hat{\rho} \\ \hat{\rho} & 1 \end{pmatrix}$$

donde $\hat{\rho}$ es el coeficiente de correlación entre las variables Y y x . Si $\lambda_1^* < \lambda_2^*$ son los autovalores de (4.15) es sabido que $\lambda_1^* = 1 - |\hat{\rho}|$ y $\lambda_2^* = 1 + |\hat{\rho}|$. Por consiguiente

$$\lambda_j^* - \lambda_j^{*(u)} = (-1)^j [|\hat{\rho}| - |\hat{\rho}(u)|] \quad (j = 1, 2) \quad (4.16)$$

Es difícil encontrar una expresión explícita de (4.16) que permita analizar qué características de un caso influyen en los autovalores de $Z^* Z^*$. Es interesante explorar los resultados que se obtienen mediante la aproximación diseñada en este apartado. Teniendo en cuenta que según (4.7), $Z^* Z^* = B[F^*_n]$, usando la fórmula (A.27) del apéndice es fácil probar que la "pseudo-curva" de influencia empírica para $\lambda_j^* (j = 1, 2)$ es

$$CIE_i(\lambda_j^*) = n \left[k_{ij}^{*2} - \lambda_j^* (\gamma_{1j}^{*2} z_i^2 + \gamma_{2j}^{*2} \gamma_i^2) \right] \quad (4.17)$$

donde $\gamma_j^* = (\gamma_{1j}^*, \gamma_{2j}^*)'$ es un autovector normalizado asociado

$$a \lambda_j^*, z_i = \frac{x_i - \bar{x}}{S_x}, \gamma_i = \frac{y_i - \bar{y}}{S_y}, k_{ij}^* = \gamma_j^{*'} z_i^* \quad y \quad z_i^* = (z_i, \gamma_i)'$$

. En este caso particular, los autovectores de $Z^* Z^*$ son

$$\gamma_1^* = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -\text{sgn}(\hat{\rho}) \end{pmatrix}, \quad \gamma_2^* = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \text{sgn}(\hat{\rho}) \end{pmatrix}$$

Para simplificar la discusión, supóngase que $\hat{\rho} > 0$. Sustituyendo en (4.17) se tiene

$$\text{CIE}_i(\lambda_j^*) = \frac{n}{2} \left[(z_i + (-1)^j \gamma_i)^2 - \lambda_j^* (z_i^2 + \gamma_i^2) \right] \quad j=1,2$$

y según (4.14),

$$\lambda_j^* - \lambda_j^*(i) = \frac{n}{2(n-1)} \left[(z_i + (-1)^j \gamma_i)^2 - \lambda_j^* (z_i^2 + \gamma_i^2) \right] + o\left(\frac{1}{n}\right) \quad j=1,2$$

de modo que si n es suficientemente grande es

$$\lambda_j^* - \lambda_j^*(i) \approx \frac{1}{2} \left[(z_i + (-1)^j \gamma_i)^2 - \lambda_j^* (z_i^2 + \gamma_i^2) \right] \quad (4.18)$$

El análisis de (4.18) es simple y depende de las coordenadas del caso $z_i^* = (z_i, \gamma_i)'$ ($i=1, \dots, n$) y de los autovalores de la matriz de correlación ampliada.

4. INFLUENCIA EN EL AUTOSISTEMA DE LA MATRIZ DE PRODUCTOS CRUZADOS

4.1. Introducción

Las cuestiones básicas que se pretenden resolver son dos:

i) dados $\lambda_1^* < \lambda_2^* < \dots < \lambda_m^*$ autovalores de la matriz de productos cruzados ampliada (tipificada o no), las diferencias

$$\lambda_j^* - \lambda_j^*(i) \quad , \quad j=1,2,\dots,m$$

¿informan de la influencia del caso i -ésimo? ¿En qué sentido? Mismas interrogantes con los autovectores.

ii) Considerando que los (λ_j^*) de la matriz de productos cruzados, o (λ_j^*) si se trabaja con la matriz ampliada, son verdaderos diagnósticos de multicolinealidad, ¿qué características de una fila de la matriz X pueden alterar o distorsionar la estructura de dependencias lineales entre las columnas de la matriz de diseño?.

La respuesta a estas cuestiones es el contenido de los apartados 4.2 y 4.3.

4.2. Influencia en los autovalores

4.2.1. Autovalores como funcionales

4.2.1.a) Datos sin tipificar

Sean $0 < \lambda_1^* < \lambda_2^* < \dots < \lambda_{m+1}^*$ los autovalores de la matriz $X^{*'}X^*$. Por consiguiente

$$\lambda_j^* = n \frac{\lambda_j^*}{n} = n T_j[F_n^*] \quad , \quad j=1, \dots, m, m+1 \quad (4.19)$$

donde

$$T_j[F_n^*] = j\text{-ésimo autovalor de la matriz } A[F_n^*] \text{ de (4.4)}$$

porque los autovalores de $X^{*'}X^*/n = A[F_n^*]$ son los de $X^{*'}X^*$ divididos por n . Para determinar la CIEi del funcional

$T_j[F_n^*]$ de (4.19) hay que usar la teoría de perturbaciones reser-
vada en el apéndice A.3.1.

Si $z^* \in \mathbb{R}^{m+1}$, se tiene

$$A[(1-\varepsilon)F_n^* + \varepsilon\delta_{z^*}] = A[F_n^*] + \varepsilon[z^*z^{*'} - A[F_n^*]] \quad (4.20)$$

y, por consiguiente, usando argumentos de continuidad y las fórmulas (A.11) el j -ésimo autovalor de (4.19) en una ordenación de menor a mayor es

$$T_j[(1-\varepsilon)F_n^* + \varepsilon\delta_{z^*}] = T_j[F_n^*] + \varepsilon[(Y_j^{*'}z^*)^2 - T_j[F_n^*]] + o(\varepsilon) \quad (4.21)$$

si ε es suficientemente pequeño. De (4.21) la CIEi para el funcional $T_j[F_n^*]$ es

$$\begin{aligned} CIE_i(\lambda_j^*) &= \lim_{\varepsilon \rightarrow 0} \frac{T_j[(1-\varepsilon)F_n^* + \varepsilon\delta_{x_i^*}] - T_j[F_n^*]}{\varepsilon} = \\ &= (Y_j^{*'} x_i^*)^2 - \frac{\lambda_j^*}{n} = (k_{ij}^*)^2 - \frac{\lambda_j^*}{n} \quad (j=1, \dots, m, m+1) \end{aligned} \quad (4.22)$$

De la expresión (4.22) y los resultados del apartado 3, se llega, finalmente, a

$$\begin{aligned} \lambda_j^* - \lambda_{j(i)}^* &= T_j[F_n^*] + (n-1)[T_j[F_n^*] - T_j[F_{n-1(i)}^*]] \approx \\ &\approx k_{ij}^{*2} \end{aligned} \quad (4.23)$$

En (4.22) y (4.23), $k_{ij}^* = Y_j^{*'} z_i^*$ es la i -ésima coordenada o

"score" de la j-ésima componente principal de la matriz X^* .

4.2.1.b) Datos tipificados

Si $\lambda_j^*(i)$ ($j=1, \dots, m$) es la notación para los autovalores de la matriz $(Z^* Z^*)_{(i)}$ no hay fórmulas explícitas relacionando λ_j^* con $\lambda_j^*(i)$ y la influencia de un caso debe estudiarse por el método aproximado del apartado 3. Es importante observar que $(Z^* Z^*)_{(i)}$ es la matriz de correlación calculada eliminando el caso i-ésimo del análisis que no es lo mismo que quitar una fila de Z^* , obtener $Z^*(i)$ y calcular $Z^*(i) Z^*(i)$.

Dado que $Z^* Z^* = B[F_n^*]$, según (4.7), λ_j^* es un funcional en F_n^* .

Es importante demostrar el siguiente resultado preliminar

LEMA 4.1.

La expresión para la perturbación $B[(1-\varepsilon)F_n^* + \varepsilon \delta_{2*}]$ es

$$B[F_n^*] + \varepsilon [I_m \otimes C[F_n^*]]^{-1/2} [(z^* - \bar{x}^*)(z^* - \bar{x}^*)' - C[F_n^*]]^{-1/2} - \frac{1}{2} C[F_n^*] \tilde{E}[F_n^*] - \frac{1}{2} \tilde{E}[F_n^*] C[F_n^*] [I_m \otimes C[F_n^*]]^{-1/2} + o(\varepsilon) \quad (4.24)$$

donde $\tilde{E}[F_n^*] = E[F_n^*] + I_m$ es una matriz diagonal de j-ésimo elemento diagonal $\frac{(z_j^* - \bar{x}_j^*)^2}{c_j[F_n^*]}$ ($j=1, \dots, m$).

DEMOSTRACION:

En la notación de la fórmula (4.6), es fácil comprobar que

$$C[(1-\varepsilon)F_n^* + \varepsilon \delta_{2*}] = C[F_n^*] + \varepsilon [(z^* - \bar{x}^*)(z^* - \bar{x}^*)' - C[F_n^*]] + o(\varepsilon) \quad (4.25)$$

donde $\bar{x}^* = (\bar{x}', \bar{y})'$. De (4.25) se sigue que

$$\begin{aligned} I_m \otimes C[(1-\varepsilon)F_n^* + \varepsilon \delta_{2*}] &= \\ &= I_m \otimes C[F_n^*] + \varepsilon I_m \otimes [(z^* - \bar{x}^*)(z^* - \bar{x}^*)' - C[F_n^*]] + o(\varepsilon) \end{aligned}$$

(\otimes = producto de Hadamard). Esta última matriz es diagonal de j-ésimo elemento

$$\begin{aligned} c_j[F_n^*] + \varepsilon [(z_j^* - \bar{x}_j^*)^2 - c_j[F_n^*]] + o(\varepsilon) &= \\ &= c_j[F_n^*] [1 + \varepsilon c_j[F_n^*] + o(\varepsilon)] \end{aligned} \quad (4.26)$$

donde $c_j [F_n^*] = \frac{(z_j^* - \bar{x}_j^*)^2}{c_j [F_n^*]} - 1$. Tomando la raíz $-\frac{1}{2}$ en el corchete de (4.26) y desarrollando en serie de Taylor de primer orden, se tiene que

$$\left[1 + \varepsilon c_j [F_n^*] + o(\varepsilon) \right]^{-1/2} = 1 - \frac{1}{2} \varepsilon c_j [F_n^*] + o(\varepsilon) \quad (4.27)$$

Si $E [F_n^*] = \text{diag} (c_1 [F_n^*], \dots, c_p [F_n^*])$, de (A.25) es

$$\begin{aligned} & \left[I_m \otimes C[(1-\varepsilon) F_n^* + \varepsilon \delta_{2*}^*] \right]^{-1/2} = \\ & = \left[I_m \otimes C[F_n^*] \right]^{-1/2} - \frac{1}{2} \varepsilon \left[I_m \otimes C[F_n^*] \right]^{-1/2} E[F_n^*] + o(\varepsilon) \end{aligned} \quad (4.28)$$

De (4.27), (4.28) y (4.7) se sigue inmediatamente la conclusión. #

De acuerdo con la fórmula (4.24) y la teoría de perturbaciones para matrices simétricas (ver apéndice A.3), la CIEi para el j-ésimo autovalor de $Z^{*1} Z^*$ es

$$\begin{aligned} CIE_i(\lambda_j^*) &= n \left[k_{ij}^{*2} - \frac{1}{n} \lambda_j^* \gamma_j^{*1} \tilde{E} [F_n^*] \gamma_j^* \right] = \\ &= n \left[k_{ij}^{*2} - \lambda_j^* \sum_{k=1}^m \gamma_{kj}^{*2} z_{ik}^{*2} \right] \quad j=1, \dots, p \end{aligned} \quad (4.29)$$

donde $k_{ij}^* = \gamma_j^{*1} z_{ij}^*$ es el "score" j-ésimo de la j-ésima componente principal de la matriz Z^* . Sea como es habitual, en (4.29) $\gamma_j^{*1} = (\gamma_{1j}^*, \dots, \gamma_{mj}^*)$ y $z_i^{*1} = (z_{i1}^*, \dots, z_{ip}^*, z_{im}^*) = (z_{i1}, \dots, z_{ip}, \gamma_i)$. En virtud de la técnica del apartado 3:

$$\lambda_j^* - \lambda_{j(i)}^* = \frac{n}{n-1} \left[k_{ij}^{*2} - \lambda_j^* \sum_{k=1}^m \gamma_{kj}^{*2} z_{ik}^{*2} \right] + o\left(\frac{1}{n}\right) \quad (4.30)$$

$j=1, 2, \dots, p, m$

4.2.2. Aplicaciones

4.2.2.a) Interpretación de resultados

La fórmula (4.23) tiene un análogo inmediato para los autovalores de la matriz $X^1 X$:

$$\lambda_j - \lambda_{j(i)} \approx k_{ij}^2 \quad j=1, 2, \dots, m \quad (4.31)$$

donde ahora $k_{ij} = \gamma_j^1 x_i$

El resultado (4.31) es particularmente atractivo por su sencillez de interpretación. Fijado un autovalor λ_j de $X^1 X$, $\lambda_{j(i)}$ es menor, en una cantidad aproximadamente igual a k_{ij}^2 .

Si $j=1$ en (4.3.1) se deduce que el módulo de multicolinealidad de Gunst desciende lo que implica que la eliminación de un caso tiende a agravar el problema de la multicolinealidad. De un modo aproximado, (4.31) informa que cuando $\gamma_j \perp x_i$, el caso no influye en λ_j . Puesto que $k_{ij} = \gamma_j' x_i = \|x_i\| \cos \theta_{ij}$, siendo θ_{ij} el ángulo que forman x_i y γ_j , la máxima influencia se consigue en las direcciones paralelas a γ_j .

Conocida la descomposición espectral de $X'X = CDC'$, los elementos de la matriz XC son los (k_{ij}) , ($i=1, \dots, n$; $j=1, 2, \dots, m, m+1$). Es sabido que las columnas de XC son ortogonales y de longitud $\lambda_j^{1/2}$ ($j=1, \dots, m, m+1$). Por tanto, en el orden de aproximación de (4.31), cuando $\lambda_j \approx 0$ ningún caso individual puede ser excesivamente influyente en λ_j . La interpretación de (4.23) es similar a la de (4.31) aunque, como ya se ha estudiado en el capítulo II, los λ_j^* ($j=1, \dots, m, m+1$) no contienen excesiva información de interés práctico.

No obstante, (4.23) tiene una importante aplicación. Sea $K^* = Z^* C^*$ la matriz de componentes principales asociada a Z^* . Si se considera la matriz $V^* = X^* (X^{*'} X^*)^{-1} X^{*'}$, se tiene:

$$V^* = X^* (X^{*'} X^*)^{-1} X^{*'} = K^* D^{*-1} K^{*'} \quad (4.32)$$

Para los elementos diagonales de (4.32) se sigue que

$$v_{ii}^* = \sum_{j=1}^{m+1} \frac{k_{ij}^{*2}}{\lambda_j^*} \quad (4.35)$$

Según (4.35), se tiene v_{ii}^* descompuesto en $m+1$ sumandos atribuible cada uno de ellos a una dirección γ_j^* . Por tanto, es natural definir el j -ésimo sumando como la influencia direccional del caso i -ésimo en la j -ésima dirección. Si

$$I_{j,i}^* = k_{ij}^{*2} / \lambda_j^* \quad j=1, \dots, m, m+1, \text{ según (4.35), es } v_{ii}^* = \sum_{j=1}^{m+1} I_{j,i}^* \text{ y despreciando el resto en (4.23) es}$$

$$\frac{\lambda_j^* - \lambda_j^*(i)}{\lambda_j^*} \approx I_{j,i}^* \quad (j=1, \dots, m, m+1) \quad (4.36)$$

Por consiguiente, de (4.36) se deduce que, en cierto sentido, la

detección de casos direccionalmente influyentes es equivalente al estudio del cambio relativo en el autosistema de la matriz X^*X^* al eliminar un caso del análisis. Puesto que basta que un caso sea direccionalmente influyente para que sea efectivamente influyente según σ_{ii}^* , la influencia de los casos en el autosistema de X^*X^* está relacionada con la detección de casos influyentes según la medida σ_{ii}^* . Obsérvese que (4.36) permite escribir

$$1 - \frac{\lambda_j^*(i)}{\lambda_j^*} \approx I_{j,i}^* \quad (j=1,2,\dots,m,m+1)$$

donde el miembro izquierdo de esta identidad aproximada depende directamente del cambio $\lambda_j^*(i)/\lambda_j^*$.

Son inmediatos los conceptos de "leverage" direccional $I_{j,i}^*$ ($j=1,\dots,m$) y la relación de éste con la influencia en el autosistema de la matriz X^*X^* .

Con datos tipificados, la interpretación de la fórmula (4.30) y su análogo

$$\lambda_j - \lambda_j(i) = \frac{n}{n-1} \left[k_{ij}^2 - \lambda_j \sum_{k=1}^p Y_{kj}^2 z_{ik}^2 \right] + o\left(\frac{1}{n}\right) \quad (j=1,\dots,p) \quad (4.37)$$

es ligeramente más complicada. Para fijar ideas, considérese en detalle la expresión (4.37) y supóngase que n es suficientemente grande. Puesto que $\sum_{k=1}^p Y_{kj}^2 = 1$, se tiene $\sum_{k=1}^p Y_{kj}^2 z_{ik}^2 \leq \max_{k=1,\dots,p} z_{ik}^2$ y por consiguiente, en la situación en la que

el "score" del caso i -ésimo en la dirección j -sima es extremo es $k_{ij}^2 > \lambda_j \sum_{k=1}^p Y_{kj}^2 z_{ik}^2$ por estar k_{ij}^2/λ_j próximo a 1. Así, en este caso, $\lambda_j - \lambda_j(i) > 0$. Sin embargo, cuando $k_{ij} = 0$ de (4.37) se deduce que $\lambda_j - \lambda_j(i) < 0$. Obsérvese que trabajando con las aproximaciones, se llega, en el contexto de datos tipificados, a que la eliminación de un caso puede aumentar o disminuir un autovalor; esto será confirmado posteriormente con otro tipo de técnicas. Finalmente, puede afirmarse con carácter general que un caso próximo al centro de los datos no puede producir una gran alteración en los (λ_j) . En efecto, si $\sigma_{ii} \approx \frac{1}{n}$ del apartado A.1 del apéndice se deduce que

$$\sigma_{ii} = \frac{1}{n} + z_{ik}^2 + m_{ik} \quad (m_{ik} \geq 0, k=1,\dots,p)$$

y de aquí $z_{ik} \approx 0$. Así pues, $k_{ij}^2 = (\gamma_j' z_i)^2 \leq \|z_i\|^2$ es aproximadamente cero y (4.37) indica que $\lambda_j - \lambda_{j(i)} \approx 0$. La interpretación de (4.30) se efectúa de manera similar.

Los "leverages" direccionales también permiten estudiar, al menos aproximadamente, la influencia en los índices de condición. Así por ejemplo, sin tipificar los datos se tiene

$$\frac{k^2[X_{(i)}]}{k^2[X]} = \frac{\lambda_{m(i)}/\lambda_{1(i)}}{\lambda_m/\lambda_1} = \frac{\lambda_{m(i)}/\lambda_m}{\lambda_{1(i)}/\lambda_1} \approx \frac{1 - I_{m,i}}{1 - I_{1,i}}$$

que depende de los "leverages" direccionales $I_{1,i}$ e $I_{m,i}$.

Ejemplo

El ejemplo de este capítulo está tomado de un trabajo de Hocking y Pendleton (1983). Se trata de un conjunto de datos artificial construido para el ajuste de un modelo

$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$. Los datos están expuesto en la tabla 4.1.

Y	x_1	x_2	x_3
57.702	12.980	0.317	9.998
59.296	14.295	2.028	6.776
56.166	15.531	5.305	2.947
55.767	15.133	4.738	4.201
51.722	15.342	7.038	2.053
60.446	17.149	5.982	-0.055
60.715	15.462	2.737	4.657
37.447	12.801	10.663	3.048
60.974	17.039	5.132	0.257
55.270	13.172	2.039	8.738
59.289	16.125	2.271	2.101
54.027	14.340	4.077	5.545
53.199	12.923	2.643	9.331
41.896	14.231	10.401	1.041
63.264	15.222	1.220	6.149
45.798	15.740	10.612	-1.691
58.699	14.958	4.815	4.111
50.086	14.125	3.193	8.453

y	x_1	x_2	x_3
48.890	16.391	9.698	-1.714
62.212	16.452	3.912	2.145
45.625	13.535	7.625	3.851
53.923	14.199	4.474	5.112
55.799	15.837	5.753	2.087
56.741	16.565	8.546	8.974
43.145	13.322	8.598	4.011
50.706	15.949	8.290	-0.248
56.817	14.123	0.578	-0.543

Tabla 4.1

i	$\frac{k_1^{*2}}{\lambda_1^*}$	$\frac{k_2^{*2}}{\lambda_2^*}$	$\frac{k_3^{*2}}{\lambda_3^*}$	$\frac{k_4^{*2}}{\lambda_4^*}$	$\frac{k_5^{*2}}{\lambda_5^*}$
5	0.0136	0.0077	0.0141	0.1319	0.0410
	0.0002	0.0209	0.0000	0.0449	0.0435
	0.0051	0.0005	0.0025	0.0010	0.0398
	0.0017	0.0001	0.0000	0.0006	0.0392
	0.0001	0.0009	0.0001	0.0176	0.0343
10	0.0561	0.0017	0.0496	0.0176	0.0460
	0.0084	0.0001	0.0099	0.0153	0.0458
	0.0771	0.0286	0.1069	0.0633	0.0192
	0.0385	0.0004	0.0599	0.0085	0.0466
	0.0113	0.0111	0.0197	0.0664	0.0380
15	0.0034	0.1677*	0.0597	0.0030	0.0437
	0.0033	0.0011	0.0037	0.0069	0.0367
	0.0150	0.0043	0.0468	0.0604	0.0354
	0.0263	0.0014	0.0299	0.0936	0.0234
	0.0116	0.0027	0.0103	0.0527	0.0492
20	0.0017	0.0012	0.0000	0.1486	0.0277
	0.0198	0.1771*	0.0010	0.0014	0.0430
	0.0304	0.4904*	0.0507	0.0295	0.0321
	0.0006	0.0225	0.0040	0.1219	0.0312
	0.0354	0.0047	0.0409	0.0002	0.0482
25	0.0304	0.0156	0.0328	0.0123	0.0270
	0.0026	0.0133	0.0027	0.0037	0.0366
	0.0057	0.0012	0.0056	0.0056	0.0394
	0.3604*	0.0124	0.1788*	0.0000	0.0419
	0.0361	0.0125	0.0631	0.0222	0.0244
	0.0002	0.0057	0.0032	0.0616	0.0332
	0.2067*	0.0036	0.2044	0.0035	0.0392

Tabla 4.2

La tabla 4.2 recoge los valores de las influencias direccionales $I_{j,i}^* = \frac{k_{ji}^*}{\lambda_j^*}$ ($i=1, \dots, 27$; $j=1, \dots, 5$) sin tipificar los datos. Destacan como puntos anómalos los casos 11, 17, 18, 24 y 27, lo que coincide con lo apuntado, por Hocking. La tabla 4.3 contiene los valores de $[\lambda_j^* - \lambda_{j(i)}^*] / \lambda_j^*$ ($i=1, \dots, 27$, $j=1, \dots, 5$).

5	0.0176	0.0033	0.0188	0.1356	0.0357
	-0.0008	0.0197	0.0000	0.0469	0.0332
	0.0036	-0.0032	0.0026	0.0010	0.0346
	0.0006	-0.0037	-0.0001	0.0005	0.0340
	-0.0009	-0.0033	0.0001	0.0181	0.0290
10	0.0592	-0.0032	0.0535	0.0177	0.0407
	0.0076	-0.0017	0.0107	0.0159	0.0405
	0.1019	0.0345	0.1223	0.0651	0.0140
	0.0398	-0.0050	0.0636	0.0084	0.0414
	0.0139	0.0093	0.0231	0.0680	0.0327
15	0.0050	0.1875*	0.0626	0.0030	0.0385
	0.0034	-0.0034	0.0038	0.0071	0.0315
	0.0189	0.0068	0.0540	0.0605	0.0302
	0.0319	-0.0016	0.0362	0.0933	0.0181
	0.0108	-0.0024	0.0177	0.0549	0.0440
20	0.0019	-0.0014	-0.0001	0.1528	0.0224
	0.0225	0.1812*	0.0011	0.0015	0.0378
	0.0810	0.5499*	0.0544	0.0292	0.0268
	-0.0003	0.0238	0.0051	0.1252	0.0260
	0.0361	0.0044	0.0431	0.0002	0.0429
25	0.0346	0.0166	0.0342	0.0122	0.0217
	0.0030	0.0108	0.0028	0.0038	0.0313
	0.0045	-0.0017	0.0059	0.0057	0.0342
	0.4610*	0.0166	0.1864*	0.0000	0.0366
	0.0426	0.0123	0.0671	0.0217	0.0192
	-0.0007	0.0037	0.0037	0.0635	0.0280
	0.2823*	0.0025	0.2138	0.0031	0.0340

Tabla 4.3

En virtud de (4.36) $\frac{\lambda_j^* - \lambda_{j(i)}^*}{\lambda_j^*} \approx I_{j,i}^*$ ($i=1, \dots, 27$, $j=1, \dots, 5$). Es de destacar la calidad de la aproximación.

4.2.2.b) Comportamiento de los gráficos de componentes principales

Hocking y Dunn (1982) desarrollan una técnica gráfica de diagnóstico basada en los gráficos de pares de componentes principales (k_j, k_ℓ) ($j, \ell = 1, \dots, p, \ell \neq j$) de la matriz K de componentes principales asociada a Z . Estos gráficos pueden tener una utilidad triple: i) detección de problemas de multicolinealidad, ii) detección de casos extremos (vii alto) y iii) detección de casos que ocultan o provocan una situación de multicolinealidad. Este tercer punto es objeto de estudio en un artículo de Gunst y Mason (1985).

La idea inicial de Hocking y Dunn es extendida en Hocking (1985) a los gráficos de componentes principales (k_j^*, k_ℓ^*) ($j, \ell = 1, \dots, m, j \neq \ell$) de la matriz K^* asociada a Z^* . El propósito de Hocking es detectar casos extremos y "outliers" con el fin de localizar casos influyentes según la medida σ_{μ}^* . Los gráficos de componentes principales son una herramienta clásica en la detección de "outliers" en análisis multivariante (Barnett y Lewis (1978)).

Sin pretender entrar en detalle en el análisis y utilización de estos gráficos, en este trabajo se desarrollan dos cuestiones no suficientemente explicadas todavía. La primera es el comportamiento de los gráficos de Hocking y Dunn ante la eliminación de un caso individual y la segunda, el comportamiento de los gráficos aumentados (k_j^*, k_ℓ^*) ante la presencia de puntos con alto leverage ampliado σ_{μ}^* . Lo más apropiado es motivar estos dos problemas con el ejemplo debido a Hocking y Pendleton (1983).

Ejemplo

La figura 4.1. muestra el gráfico (k_1, k_3) y la figura 4.2 el gráfico de las componentes (k_1^*, k_2^*)

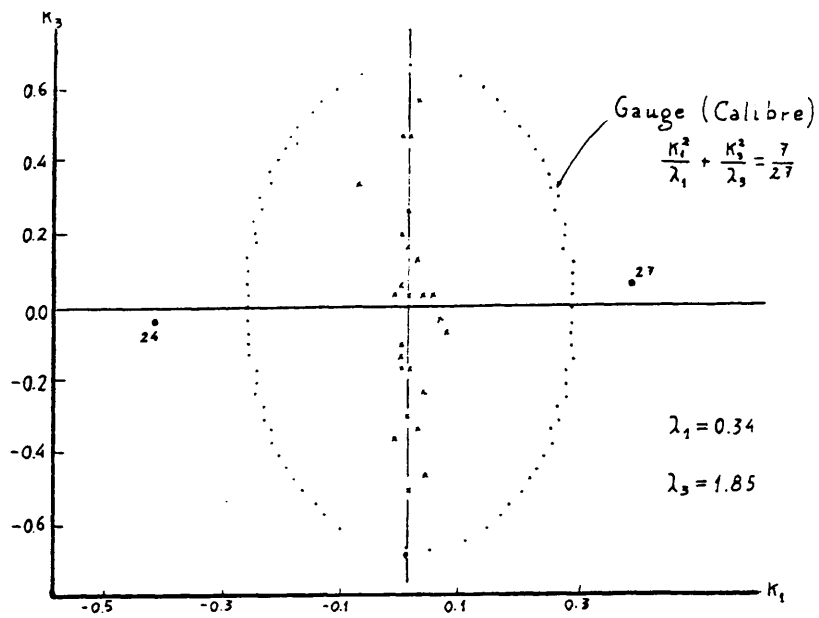


Fig 4.1

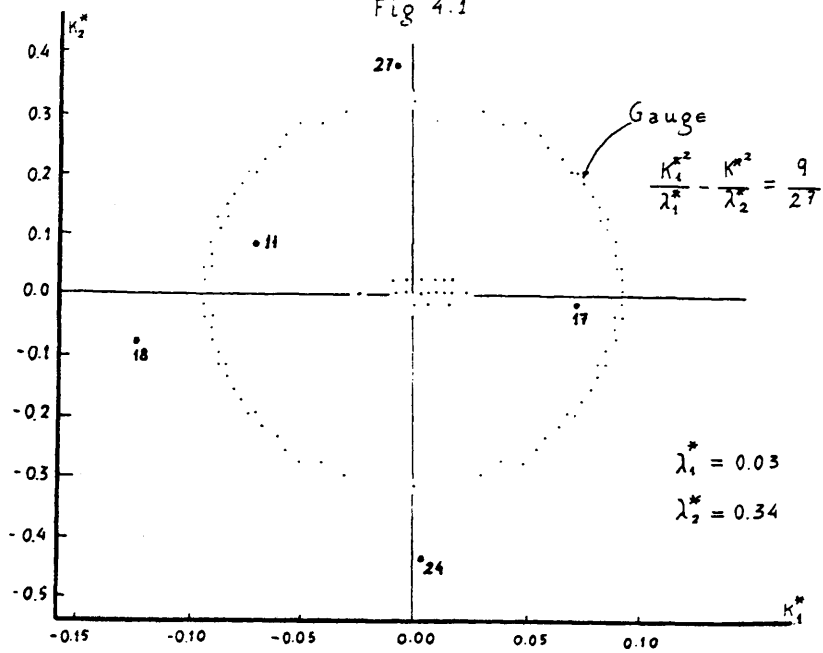


Fig 4.2

Considérese la figura 4.1. Los puntos 24 y 27 se encuentran fuera del "gauge" de modo que son, automáticamente, extremos. La anomalía de estos casos se debe a su carácter extremo en la dirección k_1 . Por otro lado, el gráfico sugiere que al eliminar los puntos 24 y 27 se producirá una situación de multicolinealidad en el modelo. Esta afirmación, de carácter intuitivo, necesita de una posterior justificación.

En la figura 4.2, los puntos 24 y 27 aparecen, de nuevo, como casos destacados. Una información nueva es la aparición de los casos 11, 18 y 17 debido al valor anómalo de los residuales. Es interesante subrayar que los puntos de alto residual se destacan en la dirección del menor de los autovalores de Z^*Z^* mientras que los puntos 24 y 27 son anómalos en la dirección de λ_1^* . Este hecho no es casual y puede explicarse de forma rigurosa empleando un resultado del capítulo II. #

La cuestión planteada en la fig. 4.1 puede contestarse con facilidad empleando la fórmula (4.37)

$$\lambda_j - \lambda_j(i) \approx k_{ij}^2 - \lambda_j \sum_{k=1}^p \gamma_{kj}^2 z_{ik}^2 \quad j=1,2,3$$

los puntos 24 y 27 son extremos en la dirección de γ_1 y, por consiguiente, su eliminación producirá un descenso en el autovalor λ_1 . Al no haber ningún punto destacado en la dirección γ_3 el autovalor λ_3 es relativamente robusto ante la eliminación de estos dos casos. Por consiguiente, los "outliers" 24 y 27 están ocultando un potencial multicolinealidad en el modelo. Esta conclusión puede obtenerse consultando los resultados numéricos de la tabla 4.4. En dicha tabla,

Casos elim.	$\lambda_1(i)$	$\hat{\lambda}_1(i)$	$\lambda_3(i)$	$\hat{\lambda}_3(i)$
I				
---	0.3400	---	1.8480	---
(24)	0.1576	0.1716	1.9873	1.9646
(24,27)	0.0177	0.0316	2.1024	2.0874

Tabla 4.4.

$\hat{\lambda}_j(i) = \lambda_j - \frac{n}{n-1} \left[k_{ij}^2 - \lambda_j \sum_{k=1}^3 Y_{kj}^2 z_{ik}^2 \right]$, $\hat{\lambda}_j(i) = \lambda_j(i) + o\left(\frac{1}{n}\right)$ según (4.37). La aproximación es suficientemente correcta, especialmente en lo que respecta a λ_3 .

Este ejemplo se puede aprovechar para desarrollar la cuestión planteada en la figura 4.2. El problema se desvía del tema general del capítulo aunque constituye una propiedad interesante de los gráficos de componentes principales. En efecto, la figura 4.2 puede analizarse en términos de una expresión conveniente de k_j^* . El desarrollo parte de la fórmula (2.25) del capítulo

$$Y_j^* = \frac{1}{[1 + u[D - \lambda_j^* I_p]^{-2} u]^{1/2}} \begin{pmatrix} -C[D - \lambda_j^* I_p]^{-1} u \\ 1 \end{pmatrix}$$

lo II. Se sabe que Y_j^* es un autovector normalizado asociado al autovalor λ_j^* de $Z^{*'} Z^*$ ($j=1, 2, \dots, m$). (La notación es la misma que en (2.25)). Por consiguiente, si $K^* = (k_1^*, \dots, k_m^*)$, es

$$\begin{aligned} k_j^* &= Z^{*'} Y_j^* = [Z | Y] Y_j^* = \frac{1}{[1 + u[D - \lambda_j^* I_p]^{-2} u]^{1/2}} [Y - K[D - \lambda_j^* I_p]^{-1} K' Y] = \\ &= \frac{1}{a_j^*} [I_n - K[D - \lambda_j^* I_p]^{-1} K'] Y \end{aligned} \quad (4.38)$$

donde $a_j^* = [1 + u[D - \lambda_j^* I_p]^{-2} u]^{1/2}$ ($j=1, \dots, m$). La fórmula de (4.38) puede reescribirse de manera más conveniente. Puesto que

$$[D - \lambda_j^* I_p]^{-1} = D^{-1} - D^{-1} \left[D^{-1} - \frac{1}{\lambda_j^*} I_p \right]^{-1} D^{-1}$$

según (A.1), haciendo las operaciones convenientes se llega a

$$[I_n - K[D - \lambda_j^* I_p]^{-1} K'] = I_n - Z(Z'Z)^{-1} Z' + \lambda_j^* M_j^*$$

donde

$$M_j^* = K Q_j^* K, \quad Q_j^* = \text{diag}(q_{1,j}^*, \dots, q_{p,j}^*)$$

siendo $q_{\ell,j}^* = [\lambda_\ell (\lambda_j^* - \lambda_\ell)]^{-1}$ ($\ell=1, \dots, p$). Por consiguiente,

$$[I_n - K[D - \lambda_j^* I_p]^{-1} K'] Y = \left[\frac{e}{S_y} + K c_j^* \right] \quad (4.39)$$

donde c_j^* es un vector $p \times 1$ de coordenadas

$$c_{j,\ell}^* = \frac{\lambda_j^*}{\lambda_\ell (\lambda_j^* - \lambda_\ell)} k_\ell' Y \quad \ell = 1, \dots, p \quad (4.40)$$

Por otro lado, la constante a_j^* de la expresión (4.38) es

$$a_j^* = \left[1 + y'K [D - \lambda_j^* I_p]^{-2} K' y \right]^{1/2} \quad (4.41)$$

Finalmente, sustituyendo (4.39) en (4.38) se llega a

$$k_j^* = \frac{1}{a_j^*} \left[\frac{e}{S_y} + K c_j^* \right] \quad j = 1, \dots, m \quad (4.42)$$

Este resultado es la base de la interpretación de los gráficos de componentes principales.

Cuando $\lambda_j^* \approx 0$, entonces las coordenadas del vector c_j^* son pequeñas, en virtud de (4.40). Por consiguiente, de (4.42)

$$k_j^* \approx \text{const} \times e$$

y, así, las componentes principales asociadas a direcciones con autovalores pequeños se comportan según el patrón de los residuales. Como $\lambda_1^* = 0.03$, este razonamiento explica la aparición de los puntos 11, 18 y 19 como destacados en la dirección de γ_1^* . En concreto, en esta situación se tiene $c_1^* = (-0.047; 0.003; 0.000)$.

Si λ_j^* es razonable, la componente de (4.42) está dominada por las coordenadas altas del vector c_j^* . En el caso del ejemplo, $\lambda_2^* = 0.344$, $c_2^* = (40.128; 0.061; 0.063)$ y, $a_2^* = 39.763$ de modo que, según (4.42)

$$k_2^* \approx \frac{e/S_y + 40.128 k_1}{39.763} = \frac{e}{1343.79} + \frac{40.128}{39.763} k_1 \approx k_1$$

Este resultado indica cómo, en el ejemplo, los "scores" de k_2^* son ligeramente superiores a los de k_1 . Esto explica lo que ocurre en el ejemplo, donde los puntos 24 y 27 están destacados en la dirección de γ_1 y, de nuevo, en la dirección de γ_2^* .

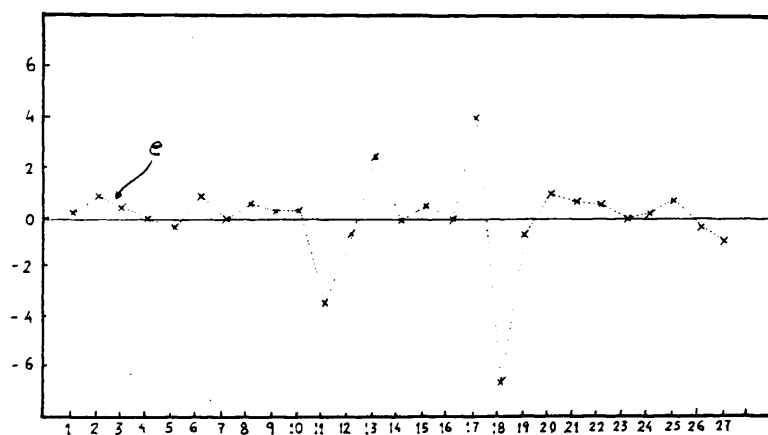


Fig. 4.3

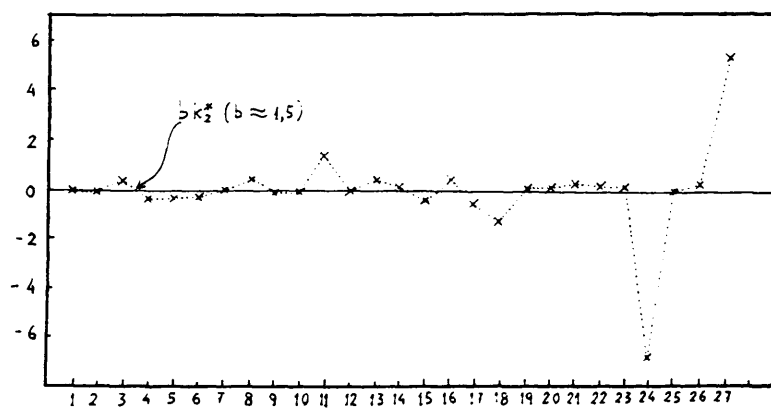


Fig. 4.4

La figura 4.3 es un index plot del vector e (trazo negro) y ak_i^* ($a \approx 51$) (trazo rojo). Por su parte, la figura 4.4 es un index plot de bk_i^* (trazo negro) y bk_i (trazo rojo) ($b \approx 15$). Es claro que $e \approx ak_i^*$ y $k_i^* \approx k_i$.

4.2.2.c) Influencia en componentes principales

Recientemente, Critchley (1985) y Joliffe (1986) han iniciado los primeros pasos en el análisis de influencia en componentes principales dentro del contexto del análisis multivariante. Uno de los problemas estudiados es la estimación, mediante la curva de influencia empírica, del cambio en los autovalores de la matriz de correlación al eliminar un caso del análisis. Joliffe en el caso de los autovalores, muestra cómo este método no es muy preciso cuando el tamaño muestral es bajo. Esta cuestión puede resolverse con la teoría desarrollada en este capítulo.

En efecto, si $Z^{*'}Z^*$ es la matriz de correlación asociada a una muestra de tamaño n de una población normal p -variante es claro que $\lambda_j^* = T_j[F_n^*]$ donde F_n^* es la f. de distribución empírica de la muestra ($j=1, \dots, p$). Por consiguiente, la técnica del apartado 3 es aplicable para llegar, de nuevo, a la fórmula (4.30). Al ser el resto de orden $O(\frac{1}{n})$, el procedimiento de estimar el cambio $\lambda_j^* - \lambda_{j(i)}^*$ usando la curva de influencia muestral es tanto más exacto cuanto mayor sea el tamaño muestral.

No es el objetivo de este trabajo entrar en detalle en el análisis de este problema. Otros artículos relacionados con la influencia en componentes principales en análisis multivariante son los de Bénasseni (1985) y Calder, Joliffe y Morgan (1986).

4.2.3. Otras técnicas

4.2.3.a) Datos sin tipificar:

El cambio en los (λ_j^*) puede juzgarse también utilizando el hecho de que al eliminar x_i^* del análisis, la matriz de productos cruzados perturbada y la original guardan entre sí la relación

$$X_{(i)}^{*'} X_{(i)}^* = X^{*'} X^* - x_i^* x_i^{*'}.$$

que es de la forma (A.10) con $C = X_{(i)}^{*'} X_{(i)}^*$, $A = X^{*'} X^*$ y $B = -x_i^{*'} x_i^*$. El único autovalor no nulo de la matriz B tiene módulo $\|x_i^*\|^2$ y está asociado al autovector x_i^* . Por consiguiente, bien aplicando el teorema de Wielandt-Hoffman de (A.17) o la desigualdad (A.18) se tiene:

$$|\lambda_j^* - \lambda_{j(i)}^*| \leq \|x_i^*\|^2 \quad j=1,2,\dots,m,m+1 \quad (4.43)$$

Esta desigualdad está de acuerdo con la aproximación dada en (4.30) porque, en virtud de la desigualdad de Cauchy-Schwartz, se tiene $k_{ij}^{*2} = [\gamma_{ij}^{*'} x_i^*]^2 \leq \|x_i^*\|^2$.

Utilizando la desigualdad (A.18) puede obtenerse mayor información. Tomando $A = X^{*'} X^*$ y $B = -x_i^{*'} x_i^*$ en (A.19) se deduce que

$$-\|x_i^*\|^2 + \lambda_j^* \leq \lambda_{j(i)}^* \leq \lambda_j^*, \quad j=1,2,\dots,m,m+1 \quad (4.44)$$

De (4.44) se deduce que los autovalores de $X_{(i)}^{*'} X_{(i)}^*$ son, uno a uno, menores o iguales que los de $X^{*'} X^*$. La mayor distancia a la que pueden estar dos autovalores correspondientes es $\|x_i^*\|^2$, el cuadrado de la norma euclídea de x_i^* .

Finalmente, puede establecerse un teorema de separación análogo al del apartado 4.2.2.a). Si en (A.20) se ponen

$$k=m+1, \quad s=2, \quad A = X_{(i)}^{*'} X_{(i)}^* \quad \text{y} \quad B = x_i^{*'} x_i^*, \quad \text{se tiene}$$

$$\lambda_{m+1-r}^* \leq \lambda_{m+1-(r-1)(i)}^* \quad r=1,2,\dots,m \quad (4.45)$$

Reuniendo (4.44) y (4.45) se tiene

$$0 < \lambda_{1(i)}^* \leq \lambda_1^* \leq \lambda_{2(i)}^* \leq \lambda_2^* \leq \dots \leq \lambda_{m(i)}^* \leq \lambda_m^* \leq \lambda_{m+1(i)}^* \leq \lambda_{m+1}^* \quad (4.46)$$

El teorema de (4.46) es un resultado de tipo global para el conjunto de autovalores de $X^{*'} X^*$ al contrario que (4.43), (4.44) y (4.45) que se refieren a un autovalor específico de $X^{*'} X^*$.

4.2.3.b) Datos tipificados

La fórmula de (4.2)

$$\hat{p}_{jk(i)}^* = \frac{\hat{p}_{jk}^* - \frac{r_{ij}^* r_{ik}^*}{n}}{\left[\left(1 - \frac{r_{ij}^{*2}}{n} \right) \left(1 - \frac{r_{ik}^{*2}}{n} \right) \right]^{1/2}} \quad \begin{matrix} j=1,\dots,m \\ k=1,\dots,m \end{matrix}$$

que es válida también cuando $j=k$ ($\hat{\rho}_{jj}=1$) es de utilidad en este contexto. Según (4.2), la matriz de correlaciones que se obtiene al suprimir un caso del análisis es

$$\begin{aligned} (Z^{*'} Z^*)_{(i)} &= E_i^{*-1} (Z^{*'} Z^*) E_i^{*-1} - E_i^{*-1} R_i^* R_i^{*'} E_i^{*-1} = \\ &= E_i^{*-1} (Z^{*'} Z^*) E_i^{*-1} - S_i^* S_i^{*'} \end{aligned} \quad (4.47)$$

donde $E_i^* = \text{diag} \left[(1 - r_{i1}^2/n)^{1/2}, \dots, (1 - r_{im}^2/n)^{1/2} \right]$ y el vector $R_i^* = \frac{1}{\sqrt{n}} (r_{i1}, \dots, r_{im})'$. La expresión de (4.47) es del tipo $C = A+B$ y puede tratarse con el tipo de técnicas descritas en el apéndice.

Es fácil ver que los autovalores de $E_i^{*-1} (Z^{*'} Z^*) E_i^{*-1}$ son los de $Z^{*'} Z^*$ relativos a E_i^{*-1} y se escribirán

$$0 < \lambda_1^* [Z^{*'} Z^* / E_i^{*-1}] < \dots < \lambda_m^* [Z^{*'} Z^* / E_i^{*-1}]$$

La condición $\lambda_1^* [Z^{*'} Z^* / E_i^{*-1}]$ se cumple con probabilidad 1 por ser $Z^{*'} Z^*$ y E_i^{*-1} dos matrices definidas positivas c.s. (Rao (1973), pág. 74).

Si $E_i^{*-1} \approx I_m$ o, de manera equivalente, si todos los $r_{ij}^2 \approx 0$ se puede poner de modo aproximado $\lambda_j^* [Z^{*'} Z^* / E_i^{*-1}] \approx \lambda_j^*$. La condición $r_{ij}^2 \approx 0$ ($j=1, \dots, m$) ocurre si $\sigma_{ii}^* \approx \frac{1}{n}$. En efecto, usando la fórmula (A.7) del apéndice es fácil probar que

$$\sigma_{ii}^* = \frac{1}{n} + Z_{ij}^{*2} + m_{ij}^* \quad (j=1, 2, \dots, m) \quad (4.48)$$

siendo $m_{ij}^* \geq 0$. Como $Z_{ij}^{*2} = (n-1/n) r_{ij}^2$, la identidad (4.48) prueba la afirmación anterior. Si algún $|r_{ij}^2| \rightarrow 1$ la aproximación $\lambda_j^* [Z^{*'} Z^* / E_i^{*-1}] \approx \lambda_j^*$ ya no será tan correcta.

De la definición de S_i^* es

$$\|S_i^*\|^2 = \sum_{j=1}^m \frac{r_{ij}^2/n}{1 - r_{ij}^2/n}$$

y, de nuevo, si $\sigma_{ii}^* \approx \frac{1}{n}$ es $\|S_i^*\|^2 \approx 0/n$ y, de modo aproximado,

$$\begin{aligned} &|\lambda_j^* [(Z^{*'} Z^*)_{(i)}] - \lambda_j^* [Z^{*'} Z^*]| \approx \\ &\approx \lambda_j^* (i) - \lambda_j^* [Z^{*'} Z^* / E_i^{*-1}] \leq \|S_i^*\|^2 \approx 0 \end{aligned} \quad (4.49)$$

De (4.49) se deduce que las grandes perturbaciones en los autovalores han de venir necesariamente de casos distantes o influyentes según σ_{ii}^* .

Pueden establecerse también las desigualdades que conducen a la obtención de un Teorema de separación análogo a (4.46). Según (A.19) y (A.20),

$$\begin{aligned} -\|S_i^*\|^2 + \lambda_j^* [Z^{*'} Z^* | E_i^{*2}] &\leq \lambda_j^* [(Z^{*'} Z^*)_{(i)}] \leq \lambda_j^* [Z^{*'} Z^* | E_i^{*2}] \\ \text{y} \quad \lambda_{m-r}^* [Z^{*'} Z^* | E_i^{*2}] &\leq \lambda_{m-(r-1)}^* [(Z^{*'} Z^*)_{(i)}] \\ &\quad r=1, 2, \dots, m-1 \end{aligned}$$

De estas dos cadenas de desigualdades se deduce que

$$\begin{aligned} 0 &< \lambda_1^* [(Z^{*'} Z^*)_{(i)}] \leq \lambda_1^* [Z^{*'} Z^* | E_i^{*2}] \leq \lambda_2^* [(Z^{*'} Z^*)_{(i)}] \leq \dots \\ &\leq \lambda_m^* [Z^{*'} Z^* | E_i^{*2}] \end{aligned} \quad (4.50)$$

Finalmente, la identidad (4.47) puede ser útil en otro contexto.

En virtud de la fórmula

$$[X'_{(i)} X_{(i)}]^{-1} = (X'X)^{-1} + \frac{(X'X)^{-1} x_i x_i' (X'X)^{-1}}{1 - \sigma_{ii}} \quad (4.51)$$

el cambio en la estimación de la varianza de una combinación lineal es positivo. Sin embargo, si se trabaja con datos tipificados, puede ser negativo. En efecto, con datos tipificados la diferencia $\text{var}[\lambda' \hat{\alpha}_{(i)}] - \text{var}[\lambda' \hat{\alpha}]$ es proporcional a $\lambda' [(Z'Z)^{-1}_{(i)} - (Z'Z)^{-1}] \lambda$. Usando el análogo para $Z'Z$ de (4.47) y el resultado de inversión adecuado es

$$\begin{aligned} (Z'Z)^{-1}_{(i)} &= E_i (Z'Z)^{-1} E_i + \frac{E_i (Z'Z)^{-1} R_i R_i' (Z'Z)^{-1} E_i}{1 - R_i' (Z'Z)^{-1} R_i} = \\ &= E_i (Z'Z)^{-1} E_i + \frac{E_i (Z'Z)^{-1} z_i z_i' (Z'Z)^{-1} E_i}{1 - \frac{1}{n} - \frac{1}{n} \sigma_{ii}} \end{aligned}$$

y, por consiguiente, la diferencia $\text{var}[\lambda' \hat{\alpha}_{(i)}] - \text{var}[\lambda' \hat{\alpha}]$ es proporcional a

$$\lambda' E_i \left[(Z'Z)^{-1} + \frac{(Z'Z)^{-1} z_i z_i' (Z'Z)^{-1}}{(1 - \frac{1}{n})^2 (\sigma_{ii} - \frac{1}{n})} - E_i (Z'Z)^{-1} E_i \right] E_i \lambda \quad (4.52)$$

Si λ es ortogonal al vector $E_i (Z'Z)^{-1}$, (4.52) se simplifica a $-\lambda' (Z'Z)^{-1} \lambda < 0$

Como conclusión, cuando la transformación depende de los datos no hay un patrón sistemático de aumento de la varianza al eliminar un caso como indica la identidad (4.51) en ausencia de tipificación.

4.3. Influencia en los autovectores

Los resultados que se obtienen con los autovectores de la matriz de productos cruzados no son excesivamente útiles y por ello el problema se trata con cierta brevedad. Para fijar ideas, el estudio se centra en los autovectores de la matriz de correlación ampliada $Z^* Z^*$. Evidentemente, un estudio paralelo podrá hacerse para las matrices $Z'Z$, $X'X$ o $X^* X^*$.

4.3.1. Autovectores como funcionales

Dado que $Z^* Z^* = B[F_n^*]$, el j-ésimo autovector Y_j^* de la matriz de correlación ampliada es un funcional en F_n^* . Según la expresión (4.24) para $B[(1-\varepsilon)F_n^* + \varepsilon \delta_2^*]$ y la fórmula (A.14) la "pseudo-curva" de influencia empírica para Y_j^* es:

$$CIE_i(Y_j^*) = \sum_{\substack{l=1 \\ l \neq j}}^m \frac{Y_l^{*'} [u^* u^{*'} - R[F_n^*] Y_j^*]}{\lambda_j^* - \lambda_l^*} Y_l^* \quad (4.53)$$

donde en (4.53) es

$$u^* = [I_m \otimes C[F_n^*]]^{1/2} (x_i^* - \bar{x}^*) = \sqrt{n} z_i^*$$

$$R[F_n^*] = \frac{1}{2} [B[F_n^*] \tilde{E}[F_n^*] + \tilde{E}[F_n^*] B[F_n^*]] \quad (4.54)$$

La expresión (4.53) se puede reescribir de manera más conveniente. Sea la matriz

$$G_j^* = \text{diag} \left(\frac{1}{\lambda_1^* - \lambda_j^*}, \dots, 0, \dots, \frac{1}{\lambda_m^* - \lambda_j^*} \right)$$

que puede expresarse también como

$$G_j^* = [D^* - \lambda_j^* I_m + c_j^* c_j^{*'}]^{-1} - c_j^* c_j^{*'} \quad (4.55)$$

donde c_j^* es el j-ésimo vector canónico de \mathbb{R}^m . Es fácil ver que (4.53) se puede poner

$$CIE_i(Y_j^*) = C^* G_j^* C^{*'} [R[F_n^*] - u^* u^{*'}] Y_j^* \quad (4.56)$$

y como, en virtud de (4.55), es

$$C^* G_j^* C^{*'} = [B[F_n^*] + Y_j^* Y_j^{*'} - \lambda_j^* I_m]^{-1} - Y_j^* Y_j^{*'}$$

sustituyendo en (4.56) se tiene, finalmente,



$$\begin{aligned} CIE_i(Y_i^*) &= n [Z^{*'} Z^* + Y_i^* Y_i^{*'} - \lambda_j^* I_m]^{-1} \cdot \\ &\cdot \left[[k_{ij}^{*2} - Y_j^{*'} \tilde{R} [F_n^*] Y_i^*] I_m - [z_i^* z_i^{*'} - \tilde{R} [F_n^*]] \right] Y_i^* = \\ &= n \tilde{CIE}_i(Y_i^*) \end{aligned} \quad (4.57)$$

En (4.57) la matriz $\tilde{R} [F_n^*]$ es

$$\tilde{R} [F_n^*] = \frac{1}{2} [(Z^{*'} Z^*) H [F_n^*] + H [F_n^*] (Z^* Z^{*'})]$$

donde $H [F_n^*]$ es una matriz diagonal de j-ésimo elemento diagonal z_{ij}^{*2} ($j=1, \dots, m$).

Según la técnica diseñada en el apartado 3, se

tiene

$$Y_i^* - Y_{j(i)}^* = \frac{1}{n-1} CIE_i(Y_i^*) + o\left(\frac{1}{n}\right) = \frac{n}{n-1} \tilde{CIE}_i(Y_i^*) + o\left(\frac{1}{n}\right) \approx \tilde{CIE}_i(Y_i^*) \quad (4.58)$$

si n es suficientemente grande. La expresión (4.58) asigna a cada caso una magnitud vectorial en \mathbb{R}^m , espacio en el que no se dispone de un orden natural. Por consiguiente, (4.58) necesita de una norma $\tilde{CIE}_i' M \tilde{CIE}_i$ para ordenar los casos según su influencia. En ausencia de otro tipo de consideraciones, una elección conveniente es tomar la matriz M definida positiva

$$M = [Z^{*'} Z^* + Y_j^* Y_j^{*'} - \lambda_j^* I_m]^{-2} \quad (4.59)$$

Tomando M como en (4.59), la expresión $\tilde{CIE}_i' M \tilde{CIE}_i$ es igual a

$$\begin{aligned} m_i^{*2} &= [k_{ij}^* z_i^* - R [F_n^*] Y_i^*]' [k_{ij}^* z_i^* - R [F_n^*] Y_i^*] - \\ &- [k_{ij}^{*2} - Y_j^{*'} R [F_n^*] Y_j^*]^2 \end{aligned} \quad (4.60)$$

Como $Y_i^* z_i^* = \sum_{k=1}^m Y_{ik}^* z_{ik}^*$, se tiene

$$Y_j^{*'} \tilde{R} [F_n^*] Y_j^* = \lambda_j^* \sum_{k=1}^m Y_{kj}^{*2} z_{ik}^{*2}$$

de manera que, según (4.60) y (4.30) se tiene

$$\begin{aligned} m_i^{*2} &\approx [k_{ij}^* z_i^* - R [F_n^*] Y_i^*]' [k_{ij}^* z_i^* - R [F_n^*] Y_i^*] - \\ &- [\lambda_j^* - \lambda_{j(i)}^*] \end{aligned}$$

resultado que, en cierto modo, relaciona la influencia de un caso en λ_j^* con la influencia en el correspondiente autovector normalizado.

4.3.2. Caso particular. (Criterio de Wilks)

Uno de los criterios más utilizados para detectar un (único) "outlier" en una muestra multivariante es debido a Wilks (1963). El método está recogido en Gnanadesikan (1977), Barnett y Lewis (1978) y Cook-Weisberg (1982). A continuación, se estudia este criterio para la detección de casos influyentes en los autovectores.

Sean c_1, \dots, c_n vectores en \mathbb{R}^m y sean también $\bar{c} = \frac{1}{n} \sum_{i=1}^n c_i$, $\Delta = \sum_{i=1}^n (c_i - \bar{c})(c_i - \bar{c})'$. El criterio de Wilks elige \bar{c}_i como posible "outlier" si minimiza la razón

$$\frac{|\Delta(i)|}{|\Delta|} \quad (4.61)$$

Como $|\Delta|$ es proporcional al cuadrado del volumen del elipsoide $\bar{c}' \Delta^{-1} \bar{c} \leq 1$, minimizar (4.61) es equivalente a minimizar el volumen restante después de eliminar $c_i - \bar{c}$ lo que, en cierto sentido, indica que c_i está alejado de los demás c_j . Usando el resultado (A.3) sobre determinantes, minimizar (4.61) es equivalente a maximizar la distancia

$$d_i = (c_i - \bar{c})' \Delta^{-1} (c_i - \bar{c}) \quad (4.62)$$

de c_i a \bar{c} relativa al elipsoide definido por Δ^{-1} .

Supóngase ahora el caso particular en el que $k_{ij}^* \approx k^*$ ($i=1, \dots, n$) para un determinado Y_j^* . Se trata de demostrar que, en esta situación, el criterio de Wilks aplicado a los $Y_j^* - Y_j^*(i)$ ($i=1, \dots, n$) es, aproximadamente, equivalente a la consideración del "leverage" σ_{ii}^* . En efecto, en virtud de (4.58), $Y_j^* - Y_j^*(i) \approx \tilde{C} E_i(Y_j^*)$ ($i=1, \dots, n$) de modo que tomando $c_i = \tilde{C} E_i(Y_j^*)$ ($i=1, \dots, n$), se tiene, en virtud de (4.57),

$$c_i' \approx \left\{ (k^2 - \alpha^2) Y_j^{*'} + Y_j^{*'} \tilde{R} [F_n^*] - k^* z_i^{*'} \right\} A^{*-1}$$

donde $A^* = Z^{*'} Z^* + Y_j^{*'} Y_j^{*-'} - \lambda_j^* I_m$ y $\alpha^2 = Y_j^{*'} \tilde{R} [F_n^*] Y_j^{*-}$. Puesto que $I_n Z^* = 0$, es

$$\bar{c}' = \frac{1}{n} \sum_{i=1}^n c_i' \approx \left[(\tilde{k}^2 - \alpha^2) Y_j^{*'} + Y_j^{*'} R [F_n^*] \right] A^{*-1}$$

y, por consiguiente,

$$(c_i - \bar{c})' \approx -k^* z_i^{*'} A^{*-1}$$

Así pues, la matriz $\Delta = \sum_{i=1}^n (c_i - \bar{c})(c_i - \bar{c})' \approx k^2 A^{*-1} Z^* Z^* A^{*-1}$ y, por tanto, la distancia d_i de (4.62) es, aproximadamente

$$d_i \approx z_i^* (Z^{*'} Z^*)^{-1} z_i^* = \sigma_{ii}^* - \frac{1}{n} = \sigma_{ii}^* [z]$$

Este resultado indica que el punto más dejado en relación con \tilde{CIE}_i corresponde al caso con mayor "leverage" σ_{ii}^* .

5. INFLUENCIA EN LOS FACTORES DE INFLACION DE LA VARIANZA

A diferencia con el autosistema existen relaciones explícitas entre los VIF_j^* y el correspondiente $VIF_j^*(i)$. Es conveniente, pues, dividir este apartado en dos secciones: una, dedicada al método directo de análisis de la relación o cambio de VIF_j^* a $VIF_j^*(i)$ y otra, dedicada a los resultados obtenidos mediante la técnica del apartado 3. Por supuesto, el estudio para los VIF_j discurre por caminos paralelos al estudio dedicado a los factores de inflación de la varianza en la matriz ampliada.

5.1. Método directo

5.1.1. Datos sin tipificar

En el contexto de datos sin tipificar interesa la expresión de la inversa de la matriz $X^{*'} X^*$. Según la fórmula de inversión (A.5) se tiene

$$(X^{*'} X^*)^{-1} = \left(\begin{array}{c|c} (X'X)^{-1} + \frac{\hat{\beta}\hat{\beta}'}{SCE} & -\frac{\hat{\beta}}{SCE} \\ \hline -\frac{\hat{\beta}'}{SCE} & \frac{1}{SCE} \end{array} \right) \quad (4.63)$$

de donde, con la notación habitual, es

$$VIF_j^* = VIF_j + \frac{\hat{\beta}_j^2}{SCE} = VIF_j \left[1 + \frac{1}{n-m} F_j \right] \quad j=1, \dots, m \quad (4.64)$$

y

$$VIF_{m+1}^* = \frac{1}{SCE} \quad (4.65)$$

Como es usual, en (4.64) F_j denota el j-ésimo estadístico F parcial.

La influencia de un caso individual en (4.65) se puede estudiar analizando los cocientes $VIF_{m+1}^*(i) / VIF_{m+1}^*$. En virtud de la fórmula citada en la introducción $SCE = SCE(i) + \frac{e_i^2}{1 - \sigma_{ii}^*}$, se

sigue que

$$\frac{VIF_{m+1}^*(i)}{VIF_{m+1}^*} = \frac{SCE}{SCE(i)} = \frac{SCE}{SCE - \frac{e_i^2}{1 - v_{ii}}} =$$

$$= \frac{1}{1 - \frac{e_i^2}{SCE(1 - v_{ii})}} = \frac{1}{1 - \frac{1}{n-m} r_i^2} \quad (4.66)$$

es una función monótona del i -ésimo residual estudentizado interno. Por consiguiente, (4.66) indica que la influencia de un caso en el último factor de inflación de la varianza de la matriz ampliada sin tipificar, es equivalente a la detección de "outliers" en el modelo (1,1).

Según (4.64), el cociente entre $VIF_j^*(i)$ y VIF_j^* ($j = 1, \dots, m$) es

$$\frac{VIF_j^*(i)}{VIF_j^*} = \left[\frac{VIF_j^*(i)}{VIF_j^*} \right] \cdot \frac{1 + \frac{1}{n-(m+1)} F_j(i)}{1 + \frac{1}{n-m} F_j} \quad j = 1, \dots, m \quad (4.67)$$

que factoriza en dos componentes diferenciadas. Para estudiar la parte relativa a los estadísticos F_j parciales, puede acudirse a la fórmula

$$F_j(i) = \frac{(n-m-1) r_i^2}{n-m-r_i^2} \cdot \frac{\left[\frac{t_j}{r_i} - \rho \left(\frac{v_{ii}}{1-v_{ii}} \right)^{1/2} \right]^2}{\left[1 + \rho^2 \frac{v_{ii}}{1-v_{ii}} \right]} \quad (4.68)$$

donde ρ es la correlación entre $\hat{\beta}_j$ y $x_i \hat{\beta}$ y t_j es el j -ésimo estadístico t-parcial (por tanto, $F_j = t_j^2$). El resultado (4.68) se debe a Cook-Weisberg (1982, pág. 112).

En relación con el segundo factor, (4.68) pone de manifiesto que la eliminación de un caso puede causar efectos muy dispares. Sin embargo, hay dos casos particulares de interés. Cuando $\rho \frac{v_{ii}}{1-v_{ii}} \approx 0$ es inmediato que si $r_i^2 > 1$

$$F_j(i) \approx F_j \frac{n-m-1}{n-m-r_i^2} > F_j$$

Por consiguiente, eliminar un caso con $v_{ii} \approx 0$ y $r_i^2 > 1$ produce un segundo factor en (4.67) mayor que 1. Si, intercambiando los papeles de caso lejano y "outlier", se supone ahora que v_{ii} es alto y r_i^2 bajo puede demostrarse que

$$F_j - F_j(i) \approx \frac{\left[\rho t_j \left(\frac{v_{ii}}{1-v_{ii}} \right)^{1/2} + r_i \right]^2}{1 + \rho^2 \frac{v_{ii}}{1-v_{ii}}} - r_i^2$$

Si $|t_j| > 1$ se sigue que $F_j - F_{j(i)} > 0$ y, así, la eliminación de un caso con $r_i^2 \leq 1$ y σ_{ii} alto produce un segundo factor en (4.67) más pequeño que 1.

En otro orden de cosas, la influencia de un caso en el segundo factor de (4.67) depende evidentemente del cambio $F_j - F_{j(i)}$. Belsley, Kuh y Welsch (1980) consideran un estadístico análogo a $F_j - F_{j(i)}$, a saber

$$DFTSTAT_{ij} = t_j - t_{j(i)}$$

para analizar la influencia de un caso individual en los estadísticos t_j individuales. Según BKW, $DFTSTAT_{ij}$ es un diagnóstico de segundo orden que se usa cuando un caso ha mostrado ser influyente por otras razones, para averiguar si los contrastes están afectados por la eliminación del mismo.

El análisis del primer factor $VIF_{j(i)} / VIF_j$ se efectúa por el procedimiento de escribir inicialmente

$$VIF_j = \frac{|(X'X)^{(j)}|}{|X'X|} \quad (j=1, \dots, m)$$

donde el superíndice (j) indica que la variable j -ésima se ha eliminado de los cálculos. En virtud de (A.3) se tiene

$$|X'_{(i)} X_{(i)}| = (1 - \sigma_{ii}) |X'X|$$

de modo que

$$\frac{VIF_{j(i)}}{VIF_j} = \frac{|X'X|}{|X'_{(i)} X_{(i)}|} \cdot \frac{|X'X^{(j)}|}{|(X'X)^{(j)}|} = \frac{1 - \sigma_{ii}^{(j)}}{1 - \sigma_{ii}} \quad (4.69)$$

donde, en (4.69), $\sigma_{ii}^{(j)}$ es el "leverage" calculado sin considerar la variable j -ésima:

$$\sigma_{ii}^{(j)} = x_i^{(j)} (X'X)^{(j)-1} x_i^{(j)}$$

Usando la fórmula (A.6) puede probarse que

$$\sigma_{ii} = \sigma_{ii}^{(j)} + q_{ij} \quad , \quad q_{ij} \geq 0 \quad (4.70)$$

y, así, (4.69) es

$$\frac{VIF_{j(i)}}{VIF_j} = 1 + \frac{q_{ij}}{1 - \sigma_{ii}}$$

El análisis depende, por tanto, de la relación entre q_{ij} y

σ_{u_i} . Si σ_{u_i} es pequeño, de (4.70) es $q_{ij} \approx 0$ de donde un caso en consonancia con los demás no puede ser excesivamente influyente en el valor de los VIF_j . Cuando σ_{u_i} sea alto no puede darse una interpretación tan sencilla del cambio (4.69).

5.1.2. Datos tipificados

Con datos tipificados, la fórmula (2.8) permite escribir el cociente

$$\frac{VIF_{m(i)}^*}{VIF_m^*} = \frac{1 - R^2}{1 - R_{(i)}^2}$$

como función de los coeficientes de determinación R^2 y $R_{(i)}^2$.

Weisberg (1980) discute el cambio de R^2 a $R_{(i)}^2$ y concluye que no hay un criterio claro de análisis. En cuanto a los

VIF_j^* ($j=1, \dots, p$), se tiene:

$$\frac{VIF_{j(i)}^*}{VIF_j^*} = \frac{VIF_{j(i)}}{VIF_j} \cdot \frac{1 + \frac{1}{n-(m+1)} F_{j(i)}}{1 + \frac{1}{n-m} F_j} \quad (4.71)$$

según (2.7). El análisis de (4.71) es similar al de (4.67) con la diferencia de que ahora el primer factor es

$$\frac{VIF_{j(i)}}{VIF_j} = \left(1 - \frac{r_{ij}^2}{n}\right) \frac{1 - \sigma_{u_i}^{(j)}}{1 - \sigma_{u_i}^2} \quad (4.72)$$

donde $r_{ij}^2 = n^2 z_{ij}^2 / \left(1 - \frac{1}{n}\right)$. De (4.6) se sigue que

$$\sigma_{u_i} = \sigma_{u_i}^{(j)} + q_{ij} = \frac{1}{n} + \left(\frac{n-1}{n^2}\right) r_{ij}^2 + m_{ij} \quad (4.73)$$

donde $q_{ij}, m_{ij} \geq 0$. Mediante (4.73) se puede analizar (4.72) de modo similar a como (4.69) era analizado mediante (4.70).

5.2. Los VIF_j^* como funcionales

Puesto que según (4.4) es $X^{*'}X^* = n A[F_n^*]$ es $(X^{*'}X^*)^{-1} = \frac{1}{n} (A[F_n^*])^{-1}$ de aquí es, para los factores de inflación de la varianza en ausencia de tipificación,

$$VIF_j^* = \frac{1}{n} T_j[F_n^*] \quad (j=1, 2, \dots, m, m+1)$$

donde $T_j[F_n^*]$ es el j-ésimo elemento diagonal de $(X^{*'}X^*)^{-1}$.

Teniendo en cuenta que $\left(A\left[(1-\varepsilon)F_n^* + \varepsilon \delta_{x_i^*}\right]\right)^{-1} = \left(A[F_n^*] + \varepsilon [x_i^* x_i^{*'} - A[F_n^*]]\right)^{-1} = (A[F_n^*])^{-1} - (A[\bar{F}_n^*])^{-1}$.

$$\begin{aligned}
 & \left[(A[F_n^*])^{-1} + D[\varepsilon] \right]^{-1} (A[F_n^*])^{-1}, \text{ donde } D[\varepsilon] = \\
 & = E [x_i^* x_i^{*'} - A[F_n^*]], \text{ la CIE}_i \text{ para el funcional } (A[F_n^*])^{-1} \text{ es} \\
 & \lim_{\varepsilon \rightarrow 0} \frac{(A[(1-\varepsilon)F_n^* + \varepsilon \delta_{x_i^*}])^{-1} - (A[F_n^*])^{-1}}{\varepsilon} = \\
 & = -(A[F_n^*])^{-1} \lim_{\varepsilon \rightarrow 0} (\varepsilon I_{m+1})^{-1} \left[(A[F_n^*])^{-1} + (D[\varepsilon])^{-1} \right]^{-1} (A[F_n^*])^{-1} = \\
 & = -(A[F_n^*])^{-1} [x_i^* x_i^{*'} - A[F_n^*]] (A[F_n^*])^{-1} = \\
 & = -\sigma^* \sigma^{*'} + (A[F_n^*])^{-1}, \quad \sigma^* = (A[F_n^*])^{-1} x_i^* \quad (4.74)
 \end{aligned}$$

y, por consiguiente, la pseudo-curva de influencia empírica para T_j^* es el j-ésimo elemento diagonal de (4.74).

Si $c_j^* = (0, \dots, \overset{(j)}{1}, \dots, 0)'$, se tiene

$$\begin{aligned}
 \text{CIE}_i (\text{VIF}_j^*) &= - [c_j^* (A[F_n^*])^{-1} x_i^*]^2 + c_j^{*'} (A[F_n^*])^{-1} c_j^* \quad (4.75) \\
 &= n \text{VIF}_j^* - n^2 [c_j^{*'} (X^{*'} X^*)^{-1} x_i^*]^2
 \end{aligned}$$

Según la técnica del apartado 3, se sigue que

$$\begin{aligned}
 \text{VIF}_j^* - \text{VIF}_{j(i)}^* &= T_j [F_n^*] + (n-1) [T_j [F_n^*] - T_j [F_{n-1(i)}^*]] = \\
 &= \text{VIF}_j^* / n + (n-1) \left[\frac{1}{n-1} \text{CIE}_i (\text{VIF}_j^*) + o(1/n) \right] = \\
 &= \left(n + \frac{1}{n} \right) \text{VIF}_j^* - n^2 [c_j^{*'} (X^{*'} X^*)^{-1} x_i^*]^2 + (n-1) o(1/n)
 \end{aligned} \quad (4.76)$$

De (4.76), despreciando el resto, el cambio relativo en los VIF_j^* es, aproximadamente

$$\frac{\text{VIF}_j^* - \text{VIF}_{j(i)}^*}{\text{VIF}_j^*} \approx \left(n + \frac{1}{n} \right) - n^2 \frac{[c_j^{*'} (X^{*'} X^*)^{-1} x_i^*]^2}{\text{VIF}_j^*} \quad (4.77)$$

De acuerdo con esta expresión, el cambio relativo en los VIF_j^* depende de un sumando que, en virtud de la desigualdad de Cauchy-Schwartz está acotado por σ_{ii}^* . Cuando σ_{ii}^* esté próximo a su cota inferior $\frac{1}{n}$, el cambio relativo está cercano a $\frac{1}{n} \approx 0$ si n es grande. Cuando σ_{ii}^* tiene un valor próximo a 1, el cambio relativo puede tomar un valor negativo.

Es interesante comparar (4.77) con las fórmulas (3.57) y (3.58) del capítulo III. Según (3.57)

$$\text{VIF}_j^* - \text{VIF}_{j(i)}^* = - \frac{1}{\text{SCE}} \cdot \frac{r_i^2 / n-m}{1 - r_i^2 / n-m} \left[\frac{\text{DFBETA}_{ij}}{r_i^2 / n-m} - \hat{\beta}_j \right]^2$$

para $j = 1, 2, \dots, m$. Por consiguiente, respuestas "outliers"

(r_i^2 alto) tienden a provocar un gran cambio en los VIF_j^* . Los casos extremos se manifestarán en la alta magnitud de $DFBETA_{ij}$. Como ya se ha indicado en el capítulo III, (3.58) es una función monótona del estadístico r_i^2 .

El estudio de influencia en los VIF_j^* puede relacionarse con la distancia de Cook $D_i \beta_j$ para la componente j -ésima del vector β que, con la notación de (4.69), es

$$D_i \beta_j = r_i^2 \frac{\sigma_{ii} - \sigma_{ii}^{(j)}}{1 - \sigma_{ii}} \quad (4.78)$$

En efecto, si se escribe la parte no estocástica de (4.78), como

$$\frac{\sigma_{ii} - \sigma_{ii}^{(j)}}{1 - \sigma_{ii}} = \frac{1 - \sigma_{ii}^{(j)}}{1 - \sigma_{ii}} - 1 = \frac{VIF_j^{(i)}}{VIF_j} - 1$$

y se despeja $r_i^2 / (n-m)$ de (3.58),

$$\frac{r_i^2}{n-m} = \frac{V_{(i)}^*}{V_{(i)}^* - 1}$$

donde $V_{(i)}^* = VIF_{m+1}^* - VIF_{m+1}^{*(i)}$, entonces

$$D_i \beta_j = (n-m) \frac{V_{(i)}^*}{V_{(i)}^* - 1} \left[\frac{VIF_j^{(i)}}{VIF_j} - 1 \right] \quad (4.79)$$

Así pues, la distancia $D_i \beta_j$ está directamente relacionada con el estudio de influencia en los factores de inflación de la varianza sin tipificar los datos.

Con datos tipificados, al ser $Z^* Z^* = B[F_n^*]$, los VIF_j^* son funcionales en F_n^* . La "pseudo-curva" de influencia empírica de VIF_j^* ($j=1, \dots, p$) es el j -ésimo elemento diagonal de la CIE_j para el funcional $(B[F_n^*])^{-1}$. En virtud de la fórmula (4.24), se tiene

$$B[(1-E)F_n^* + E\delta_{x_i^*}] = A + BD[E]B \quad (4.80)$$

donde

$$A = B[F_n^*], \quad B = [I_m \otimes C[F_n^*]]^{1/2}$$

y

$$D[E] = E[(x_i^* - \bar{x})(x_i^* - \bar{x})'] - \frac{1}{2} C[F_n^*] \check{E}[F_n^*] - \frac{1}{2} \check{E}[F_n^*] C[F_n^*] + \frac{o(E)}{E}$$

Usando (4.80), la fórmula (A.2) del apéndice y siguiendo una técnica similar al caso de datos sin tipificar, se llega a que la CIE_j para VIF_j^* ($j=1, \dots, p$) es

$$CIE_i(VIF_j^*) = n \left[z_{ij}^2 VIF_j^* - [c_j^*(Z^*Z^*)^{-1} z_i^*]^2 \right] \quad (4.81)$$

Como ya es usual, de (4.81) se sigue que

$$VIF_j^* - VIF_{j(i)}^* = \frac{n}{n-1} \left[z_{ij}^2 VIF_j^* - [c_j^*(Z^*Z^*)^{-1} z_i^*]^2 \right] + n o \left(\frac{1}{n} \right)$$

Este resultado tiene una discusión similar a (4.76).

6. OTROS DIAGNOSTICOS

Como ya ha quedado señalado en la introducción, además de los factores de inflación de la varianza y del autosistema de la matriz de productos cruzados, hay otras medidas de colinealidad que también son de utilidad práctica. Una novedad de estos diagnósticos es que existen fórmulas sencillas relacionando D con su correspondiente $D_{(i)}$.

6.1. Coeficientes de correlación

Es conocido que el coeficiente de correlación lineal entre dos variables es extremadamente sensible ante la aparición de "outliers". (Devlin et al. (1975)). Considérese la fórmula (4.2) para el coeficiente de correlación entre dos columnas z_j y z_k de la matriz de diseño tipificada $[Z]$:

$$\hat{r}_{jk(i)} = \frac{\hat{r}_{jk} - \frac{r_{ji} r_{ik}}{n}}{\left[\left(1 - \frac{r_{ji}^2}{n} \right) \left(1 - \frac{r_{ik}^2}{n} \right) \right]^{1/2}}$$

donde $r_{ij} = \sqrt{n} z_{ij} / (1 - 1/n)^{1/2}$. Usando (4.2) es fácil probar que un "outlier" marginal en cualesquiera j y k puede provocar que $\hat{r}_{jk(i)}$ sea bajo aunque \hat{r}_{jk} sea alto y puede provocar también el problema inverso, es decir, causar un valor bajo de \hat{r}_{jk} aún cuando $\hat{r}_{jk(i)}$ sea alto. Para fijar ideas, supóngase que $i=1$, y que j, k son los subíndices $j=1$ y $k=2$. Si $x_{11} = \phi_1$ y $x_{12} = \phi_2$, $(\phi_1, \phi_2 \in \mathbb{R})$, es fácil ver que

$$r_{1l} = \frac{[1 - 1/n]^{1/2} [\phi_l - \bar{x}_{l(1)}]}{[S_{l(1)}^2 + (1 - 1/n)(\phi_l - \bar{x}_{l(1)})^2]^{1/2}} \quad l=1,2$$

de donde $\lim_{\phi_l \rightarrow \pm \infty} r_{1l} = \pm 1$ ($l=1,2$). Con estos límites puede

precisarse la inconsistencia señalada de los coeficientes de correlación. Como $|r_{1l}| \rightarrow 1$ si los $|\phi_l|$ son suficientemente

grandes, el sumando dominante en

$$\hat{p} = [1 - r_{11}^2]^{1/2} [1 - r_{12}^2]^{1/2} \hat{p}_{(1)} + r_{11} r_{12}$$

es $r_{11} r_{12}$. Por tanto, un "outlier" puede hacer que \hat{p} sea alto aún cuando $\hat{p}_{(1)}$ sea bajo. Como también se cumple que

$$\frac{|r_{11} r_{12}|}{[1 - r_{11}^2]^{1/2} [1 - r_{12}^2]^{1/2}} \rightarrow \infty \text{ cuando } |\Phi_l| \rightarrow \infty \quad l=1,2$$

siempre se puede conseguir que

$$\frac{r_{11} r_{12}}{[1 - r_{11}^2]^{1/2} [1 - r_{12}^2]^{1/2}} = -\hat{p}_{(1)} + E$$

de modo que $\hat{p} = E \frac{r_{11} r_{12}}{[1 - r_{11}^2]^{1/2} [1 - r_{12}^2]^{1/2}}$ es pequeño si E es pequeño (aunque $\hat{p}_{(1)}$ sea alto).

Como $\frac{r_{ij}^2}{n} \leq \frac{\sigma_{ii} - 1/n}{1 - 1/n}$ la fórmula (4.2) también muestra que sólo pueden producirse las grandes alteraciones en el paso de \hat{p}_{jk} a $\hat{p}_{jk(i)}$ si σ_{ii} es grande.

Considerando los \hat{p}_{jk} como funcionales en F_n , no es difícil probar usando (4.27) que la CIE_i para \hat{p}_{jk} es

$$CIE_i(\hat{p}_{jk}) = n \left[z_{ij} z_{ik} - \frac{\hat{p}_{jk}}{2} (z_{ij}^2 + z_{ik}^2) \right]$$

de donde

$$\hat{p}_{jk} - \hat{p}_{jk(i)} = \frac{n}{n-1} \left[z_{ij} z_{ik} - \frac{\hat{p}_{jk}}{2} (z_{ij}^2 + z_{ik}^2) \right] + o\left(\frac{1}{n}\right)$$

y si n es suficientemente grande:

$$\hat{p}_{jk} - \hat{p}_{jk(i)} \approx \left[z_{ij} z_{ik} - \frac{\hat{p}_{jk}}{2} (z_{ij}^2 + z_{ik}^2) \right] \quad (4.82)$$

La fórmula (4.82) se puede discutir con ayuda de la relación

$$\sigma_{ii} = \frac{1}{n} + z_{ir}^2 + m_{ir} \quad (m_{ir} \geq 0), \quad r = j, k$$

de (4.73). Si $\sigma_{ii} \approx \frac{1}{n}$, todos los z_{ir}^2 son pequeños y ningún caso será excesivamente influyente en \hat{p}_{jk} . Así pues, las altas perturbaciones en \hat{p}_{jk} han de venir de la eliminación de un caso alejado del centroíde \bar{x} , que coincide con la discusión última basada en fórmulas exactas.

Finalmente, en virtud de (4.82) se tiene

$$\hat{p}_{jk} - \hat{p}_{jk(i)} \approx \left[z_{ik} z_{ij} - \frac{\hat{p}_{jk}}{2} (z_{ij}^2 + z_{ik}^2) \right]$$

Devlin et al. (1975) demuestran que desarrollando en serie de Taylor lineal alrededor del $(0,0)$ el denominador de (4.2), se llega al mismo resultado. Por consiguiente, la aproximación diseñada en la sección 3 y el enfoque exacto son, en cierto sentido equivalentes.

6.2. Determinante de $Z'Z$

El determinante de la matriz de correlaciones $Z'Z$ puede servir en ocasiones como indicativo de colinealidad entre las columnas de Z y su consideración quizá debería incluirse en cualquier estudio completo de multicolinealidad. Por ejemplo, podría servir como paso previo antes de iniciar los cálculos para los diagnósticos. Si $0 < \lambda_1 < \lambda_2 < \dots < \lambda_p$ son los autovalores de $Z'Z$, se cumple

$$|Z'Z|^{1/p} = [\lambda_1, \dots, \lambda_p]^{1/p} \leq \frac{\sum_{i=1}^p \lambda_i}{p} = \frac{\text{tr}[Z'Z]}{p} = 1$$

de manera que $|Z'Z| \leq 1$ y $|Z'Z| = 1 \iff \lambda_1 = \dots = \lambda_p = 1$ que es el ideal de ortogonalidad o perfecta condición de la matriz Z . Cuanto menor sea $|Z'Z|$, mayor es la necesidad de investigar las relaciones lineales entre las columnas de Z . El determinante de $Z'Z$ también tiene relación con el elipsoide de confianza conjunto para α en el modelo $Y = 1_n \alpha_0 + Z\alpha + E$. Dicho elipsoide está formado por todos aquellos puntos α^* de \mathbb{R}^p tales que

$$(\alpha^* - \hat{\alpha})' Z'Z (\alpha^* - \hat{\alpha}) \leq p \hat{\sigma}^2 F_{p, n-m, \alpha}$$

donde $\hat{\alpha} = (Z'Z)^{-1} Z'Y$. El volumen de este elipsoide es $\text{const} \times |Z'Z|^{1/2}$ que es grande si $|Z'Z| \approx 0$. Esta situación es otro de los problemas causados por la multicolinealidad que es el de hacer excesivamente grandes las regiones de confianza con la pérdida de precisión correspondiente. El comportamiento de $|Z'Z|$ en un ejemplo concreto, es estudiado por Mansfield y Helms (1982).

No es difícil obtener una fórmula explícita relacionando $|Z'Z|$ con $|Z'Z|_{(i)}$. En virtud de la relación

$$(1_n | Z) = X \begin{pmatrix} 1 & -\bar{x}' \\ 0 & I_p \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & S^{-1} \end{pmatrix}$$

donde $S = \text{diag}(s_1, \dots, s_p)$ ($s_j^2 = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$, $j = 1, \dots, p$)

se tiene

$$|Z'Z| = \frac{1}{n \prod_{j=1}^p s_j^2} |X'X| \quad (4.83)$$

de modo que

$$|(Z'Z)_{(i)}| = \frac{1}{(n-1) \prod_{j=1}^p s_{j(i)}^2} |X'_{(i)} X_{(i)}| \quad (4.84)$$

Se tiene $s_{j(i)}^2 / s_j^2 = 1 - r_{ij}^2 / n$ ($j=1, \dots, p$) y, por consiguiente, de (4.83) y (4.84) es

$$\frac{|(Z'Z)_{(i)}|}{|Z'Z|} = \frac{1 - \sigma_{ii}}{1 - \frac{1}{n}} \cdot \prod_{j=1}^p \left[1 - \frac{r_{ij}^2}{n} \right]^{-1} \quad (4.85)$$

Si $\sigma_{ii} \approx \frac{1}{n}$, como se sabe que $\frac{r_{ij}^2}{n} = \frac{\sigma_{ii} - \frac{1}{n}}{1 - \frac{1}{n}}$ ningún caso resultará influyente cuando tenga asociado un valor σ_{ii} bajo. Los cambios grandes han de venir con σ_{ii} alto.

En contraste con la sencillez de (4.85), la deducción de la CIE_i para $|Z'Z| = |B[F_n]|$ es bastante aparatosa. Según la expresión (4.24),

$$B[(1-\epsilon)F_n + \epsilon\delta_{x_i}] = A[\epsilon] - B[\epsilon]D^{-1}C \quad (4.86)$$

donde

$$A[\epsilon] = (1-\epsilon)B[F_n] - \frac{\epsilon}{2} [B[F_n]E[F_n] + E[F_n]B[F_n]] + o(\epsilon)$$

$$B[\epsilon] = -\epsilon u, \quad D^{-1} = (1), \quad C = u'$$

y

$$u = (I_p \otimes C[F_n^*])^{-1/2} (x_i - \bar{x})$$

En virtud de (4.86),

$$|B[(1-\epsilon)F_n + \epsilon\delta_{x_i}]| = |A[\epsilon]| [1 + \epsilon u' (A[\epsilon])^{-1} u]$$

según (A.3). Por consiguiente,

$$CIE_i(|Z'Z|) = |B[F_n]| u' (B[F_n])^{-1} u + \lim_{\epsilon \rightarrow 0} \frac{|A[\epsilon]| - |B[\epsilon]|}{\epsilon} \quad (4.87)$$

Empleando, de nuevo, (A.3) se deduce que

$$|A[\epsilon]| = (1-\epsilon)^p |B[F_n]| \left| I_p - \frac{\epsilon}{1-\epsilon} L^{-1} + o(\epsilon) \right|$$

donde $L = \frac{1}{2} [B[F_n]E[F_n] + E[F_n]B[F_n]]$. Es fácil ver que

$\left| I_p - \frac{\epsilon}{1-\epsilon} L^{-1} + o(\epsilon) \right| = P \left[\frac{\epsilon}{1-\epsilon} \right] + o(\epsilon) = g(\epsilon)$ donde P es un polinomio de grado p . Sea $q = g'(0)$. Puesto que $(1-\epsilon)^p =$

$= 1 - \varepsilon_p + o(\varepsilon)$, el límite de (4.87) es

$$|B[F_n]| \lim_{\varepsilon \rightarrow 0} \frac{(1-\varepsilon)^p g(\varepsilon) - 1}{\varepsilon} = |B[F_n]| (a-p)$$

y, de aquí es, finalmente,

$$\begin{aligned} CIE_i(|Z'Z|) &= |B[F_n]| [u' B[F_n]^{-1} u - (p-a)] = \\ &= |Z'Z| [n\sigma_{ii} - (p-a+1)] \end{aligned} \quad (4.88)$$

Esta $CIE_i(|Z'Z|)$ de (4.88) permite escribir

$$\begin{aligned} |Z'Z| - |(Z'Z)_{(i)}| &= \frac{1}{n-1} |Z'Z| [n\sigma_{ii} - (p-a+1)] + o\left(\frac{1}{n}\right) \approx \\ &\approx |Z'Z| \left[\frac{\sigma_{ii} - \frac{1}{n}}{1 - \frac{1}{n}} - \frac{(p-a)}{n-1} \right] \end{aligned} \quad (4.89)$$

Si n es suficientemente grande como para ignorar el sumando $(p-a)/(n-1)$ en (4.89) se deduce que, aproximadamente, $|(Z'Z)_{(i)}|$ desciende hasta un máximo de $|Z'Z|$ cuando $\sigma_{ii} = 1$. Por tanto, si $|Z'Z|$ es pequeño ningún caso resultará excesivamente influyente.

Finalmente, unas breves consideraciones sobre la comparación de los enfoques exacto y aproximado.

Según la técnica aproximada, de (4.89) se sigue que

$$|Z'Z| - |(Z'Z)_{(i)}| \approx |Z'Z| \left[\frac{\sigma_{ii} - \frac{1}{n}}{1 - \frac{1}{n}} - \frac{p-a}{n-1} \right]$$

de donde

$$\frac{|(Z'Z)_{(i)}|}{|Z'Z|} \approx 1 - \left[\frac{\sigma_{ii} - \frac{1}{n}}{1 - \frac{1}{n}} - \frac{p-a}{n-1} \right] = \frac{1 - \sigma_{ii}}{1 - \frac{1}{n}} + \frac{p-a}{n-1} \quad (4.90)$$

Según la técnica exacta es, de (4.85),

$$\frac{|(Z'Z)_{(i)}|}{|Z'Z|} = \frac{1 - \sigma_{ii}}{1 - \frac{1}{n}} \prod_{j=1}^p \left[1 - \frac{r_j^2}{n} \right]^{-1}$$

Cuando n es grande pues, los dos enfoques, el aproximado y el exacto, conducen a $(1 - \sigma_{ii})$ como una buena aproximación del cociente entre determinantes.

APENDICE

A.1. FORMULAS MATRICIALES

A.1.1. Fórmulas de actualización

Sea A una matriz simétrica $k \times k$ inversible y supóngase que C y B son dos matrices $k \times l$ de rango k . Si las inversas existen es:

$$[A + CB']^{-1} = A^{-1} - A^{-1}C' [I_k + BA^{-1}C']^{-1} BA^{-1} \quad (A.1)$$

Un resultado más general puede obtenerse para la inversa de la matriz $A + BDB'$, donde A y D son dos matrices inversibles de órdenes k y l , respectivamente. Se demuestra que

$$\begin{aligned} [A + BDB']^{-1} &= A^{-1} - A^{-1}B [B'A^{-1}B + D^{-1}]^{-1} B'A^{-1} = \\ &= A^{-1} - A^{-1}BE B'A^{-1} + A^{-1}BE (E + D)^{-1} E B'A^{-1} \end{aligned} \quad (A.2)$$

donde $E = (B'A^{-1}B)^{-1}$.

En relación con (A.1) y (A.2) está la fórmula que da el determinante de una matriz $M_{s \times s}$ partida en cajas

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

donde A y D son dos matrices cuadradas inversibles. Se tiene

$$|M| = |A| \cdot |D - CA^{-1}B| = |D| |A - BD^{-1}C| \quad (A.3)$$

La información sobre estas fórmulas está en Rao (1973) ((A.1), (A.2) y (A.3)) o en C-W (1982) ((A.1) y (A.3)).

A.1.2. Inversa de una matriz en cajas

Si

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \quad \begin{aligned} P &= A - BD^{-1}C \\ Q &= D - CA^{-1}B \end{aligned} \quad (A.4)$$

entonces, suponiendo que todas las inversas existen, la matriz M^{-1} se puede calcular a partir de la expresión

$$M^{-1} = \begin{pmatrix} P^{-1} & -A^{-1}BQ^{-1} \\ -D^{-1}CP^{-1} & Q^{-1} \end{pmatrix} \quad (A.5)$$

De otro modo

$$M^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BQ^{-1}CA^{-1} & -A^{-1}BQ^{-1} \\ -Q^{-1}CA^{-1} & Q^{-1} \end{pmatrix} \quad (A.6)$$

Es importante observar que si M es d.p. (en particular $C = B'$), entonces

$$\begin{pmatrix} A^{-1}BQ^{-1}B'A^{-1} & -A^{-1}BQ^{-1} \\ -Q^{-1}B'A^{-1} & Q^{-1} \end{pmatrix} = \begin{pmatrix} A^{-1}B \\ -I_r \end{pmatrix} Q^{-1} (B'A^{-1}, -I_r)$$

es una matriz d.n.n. Las fórmulas (A.5) y (A.6) pueden verse en Draper-Smith (1981).

A.2. Diferenciación de funcionales en espacios lineales

A.2.1. Definiciones y consecuencias

a) DEFINICIONES

1. Sean X, Y dos espacios normados y $F: X \rightarrow Y$ un funcional definido en un abierto $O \subset X$. F es diferenciable en $x \in O$ cuando existe una aplicación lineal y continua $L_x: X \rightarrow Y$ tal que

$$F(x+h) - F(x) = L_x(h) + o[\|h\|]$$

siendo $\lim_{\|h\| \rightarrow 0} \frac{o[\|h\|]}{\|h\|} = 0$. La expresión $L_x(h)$ es la diferencial de Fréchet (o fuerte) de F en x . El operador L_x es la derivada de Fréchet (o fuerte) de F en x . Puede probarse que si F es diferenciable en x , L_x se determina de modo único.

b) CONSECUENCIAS

1. Se demuestra, como consecuencia de la definición, que si una aplicación F es diferenciable en el sentido fuerte en $x \in X$, se tiene

$$F(x+h) - F(x) = \lim_{t \rightarrow 0} \frac{F(x+th) - F(x)}{t} + o[\|h\|] = F'(x)(h) + o[\|h\|] \quad (A.7)$$

2. Una aplicación lineal y continua es diferenciable Fréchet en todo $x \in X$ y su derivada es ella misma.

3. La diferenciabilidad Fréchet respeta las operaciones de

la estructura vectorial y cumple además la regla de la cadena:

TEOREMA (regla de la cadena)

Sean X, Y, Z espacios normados, $F: X \rightarrow Y$ y $G: Y \rightarrow Z$. Si F es diferenciable Fréchet en $x_0 \in X$ y G es diferenciable Fréchet en $y_0 = F(x_0) \in Y$, la aplicación $H: X \rightarrow Z$, $H = G \circ F$ es diferenciable Fréchet en x_0 y $H'(x_0) = G'(y_0) \circ F'(x_0)$.

A.2.2. Derivadas de orden superior

Una aplicación $F: X \rightarrow Y$ es n -veces diferenciable en $x \in X$ cuando existen n aplicaciones $F^{(k)}(x)$ ($k = 1, 2, \dots, n$) tales que $F^{(k)}(x)$ es k -lineal y continua de tal forma que

$$F(x+h) - F(x) = F'(x)[h] + \frac{1}{2!} F^{(2)}(x)[h, h] + \dots + \frac{1}{n!} F^{(n)}(x)[h, \dots, h] + o(\|h\|^n)$$

Esta información puede encontrarse más detallada en Kolmogorov-Fomin (1975), Cap. V.

A.2.3. Algunos resultados útiles

Se trata de recoger aquí los resultados que permiten garantizar la diferenciable Fréchet de los diagnósticos de multicolinealidad del capítulo IV.

A.2.3.a) Diferenciabilidad de $\int_C g(x, y) dF_n^*$

Un primer resultado de interés, es el establecimiento de la diferenciabilidad de funcionales de la forma

$$\int_C g(x, y) dF_n^* \quad (A.8)$$

donde C es un rectángulo compacto de \mathbb{R}^m y g una función continua en C (con valores en un espacio euclídeo). Para simplificar los razonamientos se supondrá que $m=1$, y por tanto, $C = [a, b]$ y $g: C \rightarrow \mathbb{R}$ con lo que la hipótesis de continuidad implica que $|g(x)| \leq M$, $\forall x \in C$. ($M > 0$).

Sean $E = \{ \alpha: \mathbb{R} \rightarrow \mathbb{R} : \alpha \text{ de variación acotada en } \mathbb{R}, \alpha(x) = 0, x \leq a, \alpha(x) = \alpha(b), x \geq b \}$

y $\|\alpha\|$ variación total de α en $[a, b]$. Es fácil probar que: i) E es un espacio vectorial sobre \mathbb{R} , ii) $\|\alpha\|$ es una

norma en E y iii) E contiene a todas las funciones de distribución en \mathbb{R} cuya medida de probabilidad está concentrada en $[a, b]$, (de hecho, E es el menor espacio de Banach que contiene a todas las funciones de distribución concentradas en $[a, b]$, Boos, Ph. D. (1977), pág. 11).

Si se define el funcional $S: E \rightarrow \mathbb{R}$ por

$$S[\alpha] = \int_a^b g(x) d\alpha(x)$$

se tiene (Apostol, 1976)

$$|S[\alpha]| = \left| \int_a^b g(x) d\alpha(x) \right| \leq M \|\alpha\|_0$$

y, por consiguiente, el funcional S , lineal en α , es continuo en E con respecto a la norma $\|\cdot\|_0$. En virtud de los resultados del apéndice, S es diferenciable Fréchet en todo $\alpha \in E$ y su derivada es el propio S .

Cuando $m \geq 2$, las modificaciones obvias conducen a la diferenciable Fréchet de (A.8) cuando α es una función de variación acotada en \mathbb{R}^m constante fuera del rectángulo $C = [a_1, b_1] \times \dots \times [a_m, b_m]$ y $g: \mathbb{R}^m \rightarrow \mathbb{R}^1$, continua, $\|g(x)\| \leq M$ en C .

A.2.3.b) Diferenciabilidad de un funcional de argumento matricial

Un resultado útil es establecer que si W es una matriz $k \times k$ de elementos diagonales no nulos, la aplicación de $\mathbb{R}^{k \times k}$ en $\mathbb{R}^{k \times k}$ definida por $W \rightarrow [I_k \otimes W]^{-1/2} W [I_k \otimes W]^{-1/2}$ es diferenciable Fréchet en todo W .

Mostrar esto no es difícil porque la diferenciable Fréchet entre espacios \mathbb{R}^q puede comprobarse componente a componente (Spivak (1975)). En este caso, la componente (i, j) es, si $W = (w_{ij})$ ($i, j = 1, \dots, k$),

$$\frac{w_{ij}}{(w_{ii} w_{jj})^{1/2}}$$

que es, evidentemente, diferenciable como función de $\mathbb{R}^{k \times k}$ en \mathbb{R} , si los elementos diagonales de W son no nulos.

A.2.3.c) Diferenciabilidad del autosistema de una matriz Simétrica

Este apartado se centra en dos resultados del li-

bro de Kato (1982) que se enuncian como teoremas.

TEOREMA A.1

Sea A una matriz $k \times k$ real y simétrica con todos sus autovalores distintos $\lambda_1[A] < \lambda_2[A] < \dots < \lambda_k[A]$. Las funciones $\lambda_j[A]$ ($j=1, \dots, k$), considerando A como argumento, son holomorfas en un entorno de A .

DEMOSTRACION:

Kato ((1982), pág. 137). #

Recuérdese que $\lambda_j[A]$ es holomorfo en un entorno de la matriz A cuando

$$\lambda_j[A+H] = \lambda_j[A] + \lambda_j^{(1)}[A, H] + \lambda_j^{(2)}[A, H] + \dots$$

donde $\lambda_j^{(n)}[A, H] = \lambda_j^{(n)}[A, H, \dots, H]$ es una forma n -lineal y continua aplicada en el complejo de matrices (H, \dots, H) .

Por tanto, la holomorfía implica la diferenciabilidad Fréchet de los autovalores como función del argumento matricial A .

Se cumple el análogo para autovectores

TEOREMA A.2

Dada una matriz real y simétrica A de orden $k \times k$ de autovalores distintos entre sí, existen k funcionales holomorfos γ_j ($j=1, \dots, k$) en un entorno de A tales que

$$\begin{pmatrix} \lambda_1[A+H] \\ \vdots \\ \lambda_k[A+H] \end{pmatrix} = [\gamma_j[A+H]]' [A+H] [\gamma_j[A+H]] \quad j=1, \dots, k$$

Los funcionales $\gamma_j[A+H]$ ($j=1, \dots, k$) son autovectores de la matriz $A+H$ y satisfacen la condición de ortogonalidad

$$\gamma_j'[A+H] \cdot \gamma_t[A+H] = \delta_{jt} \quad \begin{matrix} j=1, \dots, m \\ t=1, \dots, m \end{matrix}$$

DEMOSTRACION:

Kato ((1982), pág. 137). #

Como consecuencia inmediata de los teoremas A.1 y A.2, se tiene:

COROLARIO:

Dada una matriz real y simétrica A de orden $k \times k$ de autovalores distintos entre sí, se tiene:

- a) el funcional $A \rightarrow |A|$ es holomorfo en un entorno de A .
- b) si A es inversible, el funcional $A \rightarrow A^{-1}$ es holomorfo en un entorno de A .

A.3. Teoría de perturbaciones para matrices simétricas

Lo que sigue es una breve relación de resultados sobre los autosistemas de dos matrices reales y simétricas A y C de orden $k \times k$ que satisfacen bien una expresión del tipo

$$C[\alpha] = A + \alpha B \quad (A.9)$$

donde α es un parámetro real pequeño o bien una expresión del tipo

$$C = A + B \quad (A.10)$$

En (A.9) y (A.10), B es una matriz real y simétrica $k \times k$.

A.3.1. $C[\alpha] = A + \alpha B$

Supóngase que la matriz A tiene todos sus autovalores simples. Una herramienta analítica importante en relación con (A.9) es la llamada teoría de perturbaciones para matrices simétricas que, hablando en términos imprecisos, establece que para $|\alpha|$ suficientemente pequeño, los autovalores y autovectores de $C[\alpha]$ son series de potencias de la forma

$$\lambda_j[\alpha] = \lambda_j + \sum_{r=1}^{\infty} \lambda_j^{(r)} \alpha^r \quad (A.11)$$

$$Y_j[\alpha] = Y_j + \sum_{r=1}^{\infty} Y_j^{(r)} \alpha^r \quad (A.12)$$

donde λ_j y Y_j son, respectivamente, un autovalor y un autovector de la matriz A y los coeficientes $\lambda_j^{(r)}$ y $Y_j^{(r)}$, ($r=1,2,3,\dots$) dependen de la matriz B y del autosistema de la matriz A . En concreto, para los dos primeros coeficientes de la serie (A.4) se tiene

$$\lambda_j^{(1)} = Y_j' B Y_j$$

$$\lambda_j^{(2)} = \sum_{\substack{t=1 \\ t \neq j}}^k \frac{[Y_t' B Y_j]^2}{\lambda_j - \lambda_t} \quad j=1,2,\dots,p \quad (A.13)$$

y para los dos primeros de la serie (A.12)

$$Y_j^{(1)} = \sum_{\substack{t=1 \\ t \neq j}}^k \frac{Y_t' B Y_j}{\lambda_j - \lambda_t} Y_t \quad (A.14)$$

$$Y_j^{(2)} = \sum_{\substack{t=1 \\ t \neq j}}^k b_{t2} Y_t$$

donde, en (A.14), b_{t2} está dado por

$$b_{t2} = \frac{-[Y_j' B Y_t] \left(\frac{Y_t' B Y_j}{\lambda_j - \lambda_t} \right) + \sum_{\substack{m=1 \\ m \neq j}}^k \frac{(Y_m' B Y_t)(Y_t' B Y_m)}{\lambda_j - \lambda_m}}{\lambda_j - \lambda_t}$$

Obsérvese que tanto $Y_j^{(1)}$ como $Y_j^{(2)}$ en (A.14) son ortogonales a Y_j .

Las series de (A.11) y (A.12) convergen si $|\alpha|$ es menor que la cantidad

$$r_j = \frac{d_j}{2\alpha} \quad (j=1,\dots,k) \quad (A.15)$$

donde, en (A.15), d_j es la distancia de aislamiento del autovector λ_j de A y $\alpha = \|B\|$ = norma espectral de la matriz B .

Aplicaciones interesantes de las series (A.16) y (A.7) en problemas de física pueden verse en Courant y Hilbert (1953), Schiff (1955) o Kramers (1958). Estas series son también de utilidad en análisis numérico como puede verse en Lanczos (1957) o Wilkinson (1965). Un desarrollo sencillo y claro de la teoría de perturbaciones para el caso matricial puede verse en el libro de Carbó y Hernández (1976). Una exposición de la teoría de perturbaciones para operadores de tipo general se encuentra en el libro de Kato (1982).

A.3.2. $C = A + B$

Mientras (A.9) se refiere a un cambio pequeño en la matriz A, la expresión (A.10) se refiere a una perturbación finita de la matriz A.

En relación con (A.10) se tiene el siguiente resultado (Kato, 1982): si $\lambda_j[A] = \alpha_j$, $\lambda_j[B] = \beta_j$ y $\lambda_j[C] = \gamma_j$ (en el orden $\alpha_1 \leq \dots \leq \alpha_k$), ..., etc.), entonces si $\Phi(t)$ es una función convexa de variable real t , se tiene que

$$\sum_{j=1}^k \Phi[\gamma_j - \alpha_j] \leq \sum_{j=1}^k \Phi[\beta_j] \quad (A.16)$$

Si en particular se toma $\Phi[t] = \frac{1}{t^2}$ se tiene el teorema de Wielandt-Hoffman:

$$\sum_{j=1}^k [\gamma_j - \alpha_j]^2 \leq \sum_{j=1}^k \beta_j^2 \quad (A.17)$$

Si en (A.16) se pone $\Phi[t] = |t|^q$ ($q \geq 1$), se tiene también

$$\sum_{j=1}^k |\gamma_j - \alpha_j|^q \leq \sum_{j=1}^k |\beta_j|^q$$

Tomando la raíz q -ésima de esta expresión y considerando el límite cuando $q \rightarrow \infty$, se tiene

$$\max_j |\gamma_j - \alpha_j| \leq \max_j |\beta_j| \quad (A.18)$$

De la caracterización minimax de los autovalores (Wilkinson, 1965) se deducen las cadenas de desigualdades

$$\lambda_j[B] + \lambda_j[A] \leq \lambda_j[A+B] \leq \lambda_j[A] + \lambda_k[B] \quad (A.19)$$

$j = 1, \dots, k$

y

$$\lambda_{k-(r+s-2)}[A+B] \leq \lambda_{k-(r-1)}[A] + \lambda_{k-(s-1)}[B] \quad (A.20)$$

válida esta última si $r+s-1 \leq k$.

Las desigualdades (A.17) y (A.18) pueden verse en Lawson y Hanson (1974). La desigualdad (A.20) aparece en una nota de Anderson y Dasgupta (1963).

BIBLIOGRAFIA

1. Abramowitz, M. y Stegun, I.A. (1964). "Handbook of Mathematical Functions". Dover publications.
2. Anderson, T.W. y Dasgupta, S. (1963). Some inequalities on characteristic roots of matrices, *Biometrika*, 50, 522-524.
3. Andrews, D.F. y Pregibon, D. (1978). Finding outliers that matter. *J.R. Stat. Soc., Ser. B*, 40, 85-93.
4. Apostol, T. (1976). "Análisis Matemático". Ed. Reverté.
5. Atkinson, A.C. (1982). Regression diagnostics, transformation, and constructed variables (with discussion). *J.R. Stat. Soc., Ser. B.*, 44, 1-36.
6. Atkinson, A.C. (1986). Diagnostic Tests for Transformations. *Technometrics*, 28, 29-37.
7. Barnett, V. y Lewis, T. (1978). "Outliers in Statistical Data". J. Wiley.
8. Belsley, D.A. (1984). Demeaning conditioning diagnostics through centering. *Amer. Statistician*, 38, 73-77.
9. Belsley, D.A., Kuh, E. y Welsch, R.E. (1980). "Regression Diagnostics". J. Wiley.
10. Bénasseni, J. (1985). Influence des poids des unités statistiques sur les valeurs propres en analyse en composantes principales. *Revue de Statistiques Appliquées*, vol. XXXIII, 4, 41-55.
11. Berk, K.N. (1977). Tolerance and condition in Regression Computations. *J.A.S.A.*, 72, 863-866.
12. Boos, D.D. (1977). "The differential approach in statistical theory and robust inference". Ph. D. The Florida State University.
13. Bradley, R.A. y Srivastava, S.S. (1979). Correlation in polynomial regression. *Amer. Statistician*, 33, 11-14.
14. Calder, P., Jolliffe, I.T. y Morgan, B.J.T. (1986). Influential observations in principal components analysis: a case study. Fourth Sheffield Conference.
15. Carbó, R. y Hernández, J.A. (1976). "Introducción a la teoría de matrices". Ed. Alhambra.

16. Cook, R.D. (1977). Detection of Influential Observations in linear regression. *Technometrics*, 19, 15-18.
17. Cook, R.D. (1979). Influential observations in linear regression. *J.A.S.A.* 74, 169-174.
18. Cook, R.D. (1986). Assessment of Local Influence. *J. Roy Stat. Soc.*, vol. nº 48, nº 2, 133-169.
19. Cook, R.D. y Wang, P.C. (1983). Transformations and Influential Cases in Regression. *Technometrics* 25, 337-343.
20. Cook, R.D. y Weisberg, S. (1980). Characterizations of an Empirical influence functions for detecting influential cases in regression. *Technometrics*, 22, 495-508.
21. Cook, R.D. y Weisberg, S. (1982). "Residuals and Influence in Regression". Chapman and Hall.
22. Courant, R. y Hilbert, D. (1958). "Methods for Mathematical Physics". Interscience, 2ª ed.
23. Critchley, F. (1985). Influence in principal components analysis. *Biometrika*, 72, 627-636.
24. Cuevas, A. (1980). La curva de influencia: sus propiedades y aplicaciones. *Actas de la XII Reunión Nacional de la S.E.I.O. (Jaca)*.
25. Chatterjee, S. y Price, B. (1977). "Regression Analysis by Example". J. Wiley.
26. Devlin, S.J., Guanadesikan, R. y Kettenring, J.R. (1975). Robust estimation and outlier detection with correlation coefficients, *Biometrika*, 62, 531-546.
27. Draper, N. John, J.A. (1981). Influential observations and outliers in regression. *Technometrics*, 23, 21-26.
28. Draper, N. y Smith, H. (1981). "Applied Regression Analysis". J. Wiley, 2ª ed.
29. Farrar, D.E. y Glauber, R.R. (1967). Multicollinearity in Regression Analysis: the problem revisited. *Review of Economics and Statistics*, 49, 102-107.

30. Guanadesikan, R. (1977). "Methods for Statistical Analysis of Multivariate Data". J. Wiley.
31. Gray, J.B. y Ling, R.F. (1985). K-Clustering as a detection tool for influential subsets in regression (with discussion). *Technometrics*, 26, 305-330.
32. Graybill, F.A. (1976). "Theory and Application of the linear model". Duxbury Press.
33. Gunst, R.F. (1983). Regression Analysis with multicollinear predictor variables: definition, detection and effects. *Comm. in Statistics*, 12, 2217-2260.
34. Gunst, R.F. (1984). Toward a balanced assessment of collinearity diagnostics. Comentario a un artículo de Belsley. *Amer. Statistician.*, 38, 79-82.
35. Gunst, R.F. y Mason, R.L. (1977). Advantages of examining multicollinearity in regression analysis. *Biometrics*, 33, 249-260.
36. Gunst, R.F. y Mason, R.L. (1980). "Regression Analysis and its Application: A Data-Oriented Approach". Marcel-Dekker.
37. Gunst, R.F. y Mason, R.L. (1985). Outlier induced collinearities. *Technometrics*, 27, 401-407.
38. Hampel, F. (1974). The Influence Curve and its role in robust estimation. *J.A.S.A.*, 69, 383-393.
39. Hoaglin, D.C. y Welsch, R. (1978). The hat matrix in regression and ANOVA. *Amer. Statistician*, 32, 17-22.
40. Hocking, R.R. (1983). Developments in Linear Regression Methodology: 1959-1983 (with discussion). *Technometrics*, 219-249.
41. Hocking, R.R. (1985). Comentario a un artículo de Gray y Ling.
42. Hocking, R.R. y Dunn, M.R. (1982). Collinearity, influential data and ridge regression. Trabajo presentado en el Simposio de la Universidad de Delaware sobre Regresión Ridge.

43. Hocking, R.R. y Pendleton, O.J. (1983). The regression dilemma. *Communications in Statistics*, 12, 497-527.
44. Hoerl, A.E. y Kennard, R.W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12, 62-89.
45. Huber, P. (1981). "Robust Statistics". J. Wiley.
46. Johnson, W. y Geisser, S. (1982). Assessing the predictive influence of observations. En "Statistics and Probability: Essays in Honour of C.R. Rao". 343-358, North Holland.
47. Johnson, W. y Geisser, S. (1983). A predictive view of the detection and characterization of influential observations in regression analysis. *J.A.S.A.*, 78, 137-144.
48. Joliffe, I. (1986). "Principal Components Analysis". Springer-Verlag.
49. Kato, I. (1982). "A Short Introduction to Perturbation Theory for Linear Operators". Springer-Verlag.
50. Kolmogorov, A. y Fomin, . (1975). "Elementos de la teoría de funciones y del análisis funcional". Ed. Mir.
51. Kramers, M.A. (1958). "Quantum mechanics". North Holland.
52. Lanzcos, C. (1957). "Applied Analysis". Prentice-Hall.
53. Lawson, Ch.L. y Hanson, R.J. (1974). "Solving Least Squares Problems". Prentice-Hall.
54. Little, J.K. (1985). Influence and a Quadratic form in the Andrews and Pregibon Statistic. *Technometrics*, 27, 13-15.
55. Mansfield, E.R. y Helms, B.P. (1982). Detecting multicollinearity. *Amer. Statistician*, 36, 158-160.
56. Marquardt, D.W. (1970). Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation. *Technometrics*, 12, 591-612.
57. Marquardt, D.W. y Snee, R.D. (1975). Ridge Regression in practice. *Amer. Statistician*, 29, 3-20.
58. Mickey, M.R., Dunn, O.J. y Clark, V. (1967). Note on the use of stepwise regression in detecting outliers. *Computers and Biomedical Research*. 1, 105-112.

59. Patnaik, P.B. (1949). The noncentral chi-squared distributions and F distributions and their applications. *Biometrika*, 36, 202-232.
60. Radhakrishnan, R. y Kshirsagar, A.M. (1981). Influence functions for certain parameters in multivariate analysis. *Communications in Statistics*, 10, 515-529.
61. Raiffa, H. y Schlaiffer, R. (1961). "Applied Statistical Decision Theory". Harvard University.
62. Rao, C.R. (1973). "Linear Statistical Inference and its Applications". J. Wiley, 2ª ed.
63. Río, M. del (1986). Un proceso de ortogonalización para el modelo lineal. *Trabajos de Estadística*, vol. I, nº 1, 112-126.
64. Schiff, L.J. (1955). "Quantum mechanics". McGraw-Hill.
65. Seber, G.A.F. (1977). "Linear Regression Analysis". J. Wiley.
66. Silvey, R.D. (1969). Multicollinearity and imprecise estimation. *J.R. Stat. Soc., Ser. B*, 31, 539-552.
67. Spivak, M. (1975). "Cálculo en variedades". Ed. Reverté.
68. Stewart, G.W. (1973). "Introduction to Matrix Computations". Ac. Press.
69. Webster, J.T., Gunst, R.F. y Mason, R.K. (1974). Latent root regression analysis. *Technometrics*, 16, 513-522.
70. Weisberg, S. (1980). *Applied Linear Regression*. J. Wiley (1ª ed.).
71. Weisberg, S. (1983). Some principles for regression diagnostics and influence analysis. Comments on "Developments in Linear Regression Methodology: 1959-1983". Universidad de Minnesota. Tech. Report nº 416.
72. Weisberg, S. (1985). *Applied Linear Regression*. J. Wiley, (2ª ed.).
73. Wilkinson, J.H. (1965). "The Algebraic Eigenvalue Problem". Oxford University Press.
74. Wilks, S. (1963). Multivariate Statistical Outliers. *Sankhya A*, 25, 507-526.

75. Willan, A.R. y Watts, D.G. (1978). Meaningful collinearity measures. *Technometrics*, 20, 407-412.

